DOCUMENT RESUME

ED 040 726

LI 002 046

TITLE	EUDISED: European Documentation and Information System for Education. Volume III, Technical Studies.
INSTITUTION	Council of Europe, Strasbourg (France). Documentation Center for Education in Europe.
PUB DATE	69
NOTE	120p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$6.10
DESCRIPTORS	Automation, *Computer Oriented Programs, Documentation, *Education, *Foreign Countries, *Information Centers, *Information Processing, Information Systems, International Organizations, International Programs, Library Technical Processes
IDENTIFIERS	*Documentation Centre for Education in Europe

ABSTRACT

An information system, either manual or computerized, is a system for collecting, selecting, validating and augmenting data to form a collection of information (the data base) which will form the basis from which various types of information services can be provided. These services may be individual (retrospective searching or current awareness) or of a more general nature (production of catalogues, indexes, bulletins, abstract journals, etc.) In this report the technical aspects of developing an information system on education are explored in the following sections: (1) compatibility of computer systems, (2) problems of thesaurus construction for education, (3) computer handling of social science terms and their relationships, (4) cost-benefit analysis procedures and applications, (5) cost estimates for bibliographical searching in a social science information system and (6) electronic recording of educational data in the Canton of Geneva. Brief resumes on the author of each section are appended. (NH)

ED040726

COUNCIL OF EUROPE

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EUDISED

EUROPEAN DOCUMENTATION AND INFORMATION SYSTEM FOR EDUCATION

Volume III

TECHNICAL STUDIES

LI 002 046

DOCUMENTATION CENTRE FOR EDUCATION IN EUROPE

Editor :
Director of Education
and of Cultural and Scientific Affairs
Council of Europe
STRASBOURG
1969

CONTENTS

Page

A G Price
Compatibility of Computer Systems 5

H H Sträter
Problems of Thesaurus Construction for Education 23

E J Coates
Computer Handling of Social Science Terms and their Relationships 53

M H Peston
Cost-Benefit Analysis: Procedures and Applications 85

G K Thompson
Cost Estimates for Bibliographical Searching in a Social Science Information System 101

W Hutmacher
Electronic Recording of Educational Data in the Canton of Geneva 111

Notes on Authors 127

COMPATIBILITY OF COMPUTER SYSTEMS

by

A G PRICE

Introduction 7

Structure of a typical self-contained computer based information processing system 7

Characteristics of a computer readable data base 9

Data selection 13

Information processing systems 14

Conclusion 17

Appendix. International and other standards relevant to compatibility 19

Bibliography 21

1. INTRODUCTION

A computer based information system is a system for collecting, selecting, validating and augmenting data to form a collection of information (the data base) which will form the basis from which various types of information service can be provided. These may be individual services, such as retrospective searching, or current awareness, or may be of a more general nature, such as the production of catalogues, indexes, bulletins, abstract journals, or the like. A considerable proportion of the work involved is of a clerical nature, involving the selection, reformatting and resequencing of the data, and as such is admirably suited for performance by a computer. In this report we shall be concerned only with systems in which a large proportion of the essentially clerical work is performed by computer.

Apart from the service which the system was designed to provide, two tangible assets will result from the establishment of the system. One is the accumulated data and the other is the set of computer programs and the organisation necessary to perform the appropriate data processing. In view of the high cost of preparing a data base and the systems to process it, there would be an obvious advantage in securing as wide a use of each of these as possible. It is therefore the purpose of this report to study the problems involved in making use of data bases and computer information processing systems beyond that originally envisaged for them.

2. STRUCTURE OF A TYPICAL SELF CONTAINED COMPUTER BASED INFORMATION PROCESSING SYSTEM

The majority of present computer based systems provide information services based on titles, abstracts, or indexed references to articles in scientific periodicals. Data serving as input to the system is therefore obtained by means of a team of suitably qualified staff who read those journals which are expected to provide input. To ensure wide coverage it is necessary to scan many journals with a low proportion of articles relevant to the speciality of the system. When an article has been identified as relevant to the system, it is then prepared for input to the computer process. This will involve identifying those elements of information which the system needs to store, such as the title, author, bibliographic reference and an abstract or keywords which will need to be generated by a suitably qualified person. All this information is presented in an agreed form and is then put into a form suitable for computer input. This may involve retyping the material for optical character reading, or preparation of the material on paper tape or punched cards, or directly onto magnetic tape, as a computer is not capable of reading normal printed text owing to the variety of formats and type faces in use. The activity of obtaining, selecting and preparing data up to the point where it is in a form suitable for computer input may well occupy a team of twenty to fifty people full time for a single scientific discipline and it can therefore be seen to be an expensive process.

The first stage of computer processing will normally include some form of checking to ensure that, where possible, inaccurate data is not introduced into the system. Checking may consist of performing consistency checks on each item of data, for example, that the language of origin is consistent with the country of origin. Checking may also be performed against authority lists to ensure, for instance, that the journal title is accurate by comparison against a master list of all the journals which are scanned, and that keywords or indexing phrases which have been used are acceptable; this topic will be expanded later. The computer at this stage cannot check that the indexing terms or abstract which have been produced for the original article do in fact accurately represent its content, and it is therefore normally necessary to introduce some form of manual system for ensuring adequate quality control on input. It should be noted at this stage that, by the time the data reaches computer processing stages, it has already been selected from the point of view of the scientific discipline of interest to the system, and to the purposes to which the data is to be put. When we consider that of the original journal articles less than 10% by volume of the text will be incorporated in the computer system, it is obvious that some fairly extensive selection and compression process must take place. This can create difficulties, especially with articles of interest to more than one specific scientific discipline, as it is difficult to abstract or index these articles in such a way that the abstract or index terms can adequately represent the interest in the article for a multitude of scientific disciplines, unless the abstract itself is as long as the article.

When the data reaches the computer, it is converted on an initial run, which may well be part of the validation process, into a form most suitable for the type of computer in use and the processes which are subsequently to be performed on it. These processes may include the reduction of text to coded form either by statistical processes or by replacement of text with codes derived from a master authority list, such as a thesaurus of index terms. This is the method used in, for example, the National Library of Medicine MEDLARS System. In other cases the data is preserved in the form of text, as in the American Chemical Society Chemical Information Retrieval System.

If it is required to provide services based on retrospective examination of data, it will now be necessary to incorporate the new material with the old data that has been accumulated during the life of the system. On completion of this process the data base is in a form suitable for the computer production of the information services desired.

The basis of an information system is that it shall select from the totality of data available that which is considered relevant for the application concerned, and that the data shall then be presented in a suitable sequence and in a suitable format. The selection process will depend to a large extent on the use to which the data is to be put. In some cases all the data may be required as part of the output, for instance, in the production of author lists or catalogues, or abstract publications intended for a wide audience. In other cases, especially for services offered to individuals, it will be necessary to permit very complex processes of selection ranging from statistical processes of counting the similarity between terms and phrases specified by the user and those occurring in the data, through to systems involving logical combinations or exclusions of terms selected by the user from a fixed list. With existing computer based systems the formulation of the selection process for an individual user is a highly skilled activity and is best performed by someone with experience of the system. Requests involving twenty or thirty selection criteria are common and, in some instances, especially in current awareness systems where the user's need can be refined over a long period, queries involving over a hundred search terms are not uncommon. The system must therefore contain a program capable of accepting the specification of a selection process and of using it to select from the data those terms which satisfy the criteria. What constitutes satisfaction of the criteria is a subject meriting a report in its own right, and may vary from a numeric measure of the likelihood that any particular item satisfies the expressed need to a logical decision of relevance or non-relevance leading to inclusion or exclusion of any specific item.

To facilitate use of the output, especially in the case of general purpose output intended for a large number of users, it is often desirable to be able to impose some form of order on it. In the case of bibliographic references, this may involve grouping them by author in alphabetical order of authors' surnames, or by journal, by institution of origin, by some form of classification of the subject matter, or indeed by any other criteria which can be expressed in terms of constituents of the data. It will be seen that this is in many respects similar to the process of catalogue production, and indeed many different types of information service present similar types of processing problem. Finally, for human use, it is appropriate to consider the format in which the results are to be presented, as good presentation of results can make a significant difference to their usefulness. Output formats may range from listing on ordinary paper to the production of individual catalogue cards to permit the user to select and retain those items in which he is interested. In some instances a user is interested in retrieving information in a form suitable for inclusion in his own local computer based information system, and may require output in the form of punched paper tape, punched cards or magnetic tape. For the production of abstract journals where a large number of copies of the output are required, output may be produced direct on microfilm, or fed to a typesetting machine ready to produce copies via conventional printing processes. Finally, the results are disseminated to the users of the system.

It will be seen that our typical system falls fairly conveniently into two parts: the first is the collection of data and its transformation into a form suitable for computer processing, and the second is the execution of various computer based processes on the data to produce the forms of output required. The question of interest is whether the data collection and preparation process can serve to supply data to the retrieval and output processes of a different system. In the next section we shall therefore concentrate on the data base and study its characteristics to determine to what extent it is likely to be generally useful.

3. CHARACTERISTICS OF A COMPUTER READABLE DATA BASE

For a data base to be readable by computer and useful as a basis for processing, its physical characteristics must be such that it can readily be accessed by the computer, and in general it is necessary also that it should be capable of being accessed at a sufficiently high speed to take advantage of the processing power of the computer. It must also be in a well defined form, since the computer has no intrinsic powers of deduction and requires that it be presented with material expressed in a formal and precise way.

3.1. *Physical storage media for data bases*

In general one can expect that the data will be processed several times by computer, either as the basis for different types of service, or as part of a cumulative exercise where the growing data base is repeatedly processed over a long period of time. On the grounds therefore of ease of storage and speed of access, the data base is normally stored on some form of magnetic medium, such as magnetic tape, disc files, or magnetic cards. These all have the advantage of relative permanence, ease of change, high storage capacity and ready accessibility by the computer. There is, however, an essential difference between magnetic tape storage and storage on other types of magnetic device. This is that magnetic tapes have been used for storage of large volumes of data for a longer period of time than all other media. Magnetic tapes are serially accessible only, i.e. it is necessary to process the tape consecutively from beginning to end to find any particular item. There is therefore a strong economic incentive to process all the data on the tape in the sequence in which it occurs. This restriction does not apply to discs and magnetic card files. In addition a magnetic tape file can be theoretically of indefinite length, as data can be continued from one reel of tape to another and files of data occupying twenty, forty or fifty reels of magnetic tape are not uncommon. Magnetic tape represents a relatively cheap storage medium approximately ten times cheaper than magnetic discs and cheaper by a smaller factor than magnetic card files. There is also a considerable degree of compatibility between different computers with regard to their ability to handle magnetic tapes generated by another computer.

Disc files and magnetic card files, on the other hand, have the property that it is not necessary to read all the data from the beginning each time, in other words there is a random access capability, and in general systems which use discs or magnetic cards for storage take advantage of this facility. If a system depends on random access capability, then obviously all of the data base must be available to the computer simultaneously, and this effectively limits the maximum size of data base which can be processed. The logical structure of the data stored on random access devices is also, for reasons of efficiency, closely related to the physical characteristics of the storage medium and, although this may increase the efficiency and reduce the costs of the processing system, it renders the possibilities of compatibility with the similar characteristics of another processing system remote. In many instances the storage medium itself is not actually removable from the computer, as opposed to magnetic tapes which can readily be removed and taken away.

In view therefore of the physical compatibility across a wide range of machines, mobility and low relative cost of magnetic tape, it is commonly accepted as the only practicable medium by which data can be interchanged between computer based systems. Other media such as paper tape, or punched cards, suffer from the disadvantage that they are bulky, their capacity for data storage per unit volume is low, and the cost of storage per element of data is surprisingly high. In addition there are a variety of conventions for the storage of data on paper tape and punched cards, so that the probability that data prepared for one computer will be acceptable without modification to another is small. In practice, therefore, for present systems and for the foreseeable future magnetic tape may be expected to remain the main medium on which data is interchanged between computer systems. With the existence of standards for the physical properties of magnetic tape and for the handling of data stored on magnetic tape there is now no reason why almost any computer system should not be capable of generating data in such a form as it can be processed successfully by another computer. Physical compatibility, however, although obviously a prerequisite, is by no means the only or indeed the largest problem to be faced in interchanging data between systems.

3.2. *Logical structure of a data base*

Any data base needs some form of logical structure, whether the data base is computer processable or not. Consider, for instance, the Scientific Journal; the reader may be interested in an article of a specific title, or by a specific author but, having identified the page containing the title or author, he then needs a convention to enable him to decide which text is associated with that title or author, he needs to know where to start reading and, at the lowest level, he needs a convention to enable him to determine when he has read any one word which word to read next. The normal convention is that, having read any one word, the succeeding word is that which falls immediately to its right, except that if there are no more words on the line, the next word is that which begins the line following next down the page. This may seem fairly trivial but one should bear in mind that in certain languages, such as Arabic or Chinese, this convention does not hold. In the case of a catalogue entry this will describe one journal article, for instance, by giving its title, author, the journal name, the author's affiliation, the date, pagination and possibly cataloguing information such as keywords. By implication all the data on one card refers to one entity, and we know that when we move to the next card, unless a continuation convention has been established, all the information on it refers to a different bibliographic entity. In a similar way before the data stored in a computer readable data base can be made use of, it is essential that information about the logical structure of that data is available. Some part of the logical structure must be known to the processing program, or more specifically must have been known to the programmer who wrote the processing program. The program may be written to take advantage of the known or expected structure of the data, or alternatively, it may be written in such a way that the logical structure of the data can be described independently, as is the case, for instance, if the program is written in the high level language COBOL. This form of low level structure which would be reasonably apparent if a direct transcription of the data to human readable form were performed is referred to as the Explicit Logical Structure. Explicit structure is normally of a fairly simple type where data elements form a simple hierarchy or tree, as, for instance, a number of authors often relate to the same title the joint authors of the paper, or a number of titles, all of which relate to articles published in the same issue of a journal. Information which relates to the same heading information is linked in a type of family tree, and the fewer steps up and down the tree it is necessary to move to proceed from one data item to another, then the more closely related those two data items are. Even at such a simple level logical structures tend to be defined for particular computer information processing systems to take particular advantage of the characteristics of the system and known or expected characteristics of the data. These conventions may make it difficult to transfer data to a different system with different conventions, and they may also make it difficult for the processing programs which take advantage of restrictions in the explicit structure to process data which requires more generality, or even merely a different set of conventions. Considerable interest has been expressed in this problem, with particular reference to the interchange of scientific reference information; for instance, by the American standards organisation (USASI) and by the United Kingdom standards organisation (BSI). One result of this interest is a proposal for a general purpose explicit data structure, which is capable of accommodating the wide variety of types of data at the expense of some increased processing requirement to pack and unpack data into and out of the structure. This approach has been taken because it is apparent that it is already too late to define specific structures and expect users to adhere to them. The current approach therefore is more on the lines of defining methods of describing data structures which can be used as input to a computer to control the processing of data supplied in the format described.

In many instances the explicit structure used reflects constraints imposed by the environment in which the system operates. Thus, for instance, if input is from English language journals only, it will not be considered necessary to make provision to indicate language of origin. If the system is designed to handle articles from scientific journals, it may not be capable of handling the different kind of information required to identify material published in book form. All these limitations, constraints and special cases tend to be allowed to influence the design of data structure for individual computer based information processing systems to facilitate efficient processing, but they render the probability of either the data or the processing system being useful for other applications very small. It is in the majority of cases true to say that if systems and their data bases are to be made more generally available they must be planned with this object in view right from the start, as the expense of transforming data bases and systems subsequently is comparable with that of writing them again from the start.

A significant problem in construction of a data base is the recording of data in such a way that the relationships expressed on the printed page by juxtaposition, the use of variations in type face, punctuation etc. can be expressed in such a way that they are meaningful to the computer; the method used, for instance, in the proposed USASI standard for information interchange relies on a human operator assigning to each element of the data base before it is prepared for computer input a tag, or identifying number, which describes what kind of element follows, and also the relationship which this element holds to the others in the vicinity. This is a time consuming and expensive process and, in the conversion of large files to computer readable form, may involve an unacceptable expense. Another approach which has been tried on an experimental basis, using sample data from the British Museum catalogue of printed books is to describe the layout of a bibliographic reference in terms of a formal grammar or syntax. Where the data base, although conforming to a complex structure, nevertheless follows a well-defined pattern, this approach shows promise of permitting the computer to play a much larger part in identifying constituents of the data base and arranging them in a way convenient for subsequent processing. This work is being carried on at the Documentation Processing Centre of the United Kingdom Department of Education and Science. To facilitate the use of computer programs for providing the same type of processing for different forms of input, an alternative approach is for the data itself to carry a description of its own format in a form suitable for computer input. A method developed by Peter King, late of Aberystwyth University and now at Birkbeck College, London, involves associating with the data a small collection of computer routines which will make available corresponding elements of the data on demand. A processing program can then be written to perform the required functions regardless of the form of the input data, as by calling the appropriate routine, it can obtain access to any data element which it requires. To process a different but compatible data base, it is necessary to change only the access routines and not the whole processing problem.

It is therefore fairly clear that the problems of processing data bases on computers other than the ones for which they were originally designed can be carried out as long as the information is of the same kind, and that problems of differences in format, layout, storage medium and so forth, are relatively insignificant from the intellectual point of view, although their mechanical implementation may be tedious and expensive.

3.3. *Implicit data base structure*

So far we have covered the problem of identifying particular data elements for processing purposes. However, from the computer point of view, the processing program does not know in any real sense what is the significance to the human being of the information in which it is operating; a computer may be requested to search for papers by a certain author, or for papers which have been indexed by a certain term; the processing program does not know the difference between an author and an indexing term and, in fact, it has to be instructed to search for data items containing a certain value in a certain field where the human being knows that the specified field or data element represents an author name or an indexing term. A human being can generally expect to be able to recognise the particular type of a data element from its content, although in some cases even this will break down. For instance:

"Compatibility of computer systems.   A. G. Price"

is fairly obviously a title followed by an author name, whereas

"Sherlock Holmes.   Sir Arthur Conan Doyle"

is an example of the same kind as the previous one in which rather more knowledge of the context is needed to determine which represents the title and which represents the author. In the traditional citation a great deal of reliance is placed on such textual clues to assist the human reader to identify the meaning of what he is reading. In addition, especially for catalogued or classified information, although each bibliographic reference appears to stand alone and is self-contained, in practice the classification scheme provides an overall framework within which each reference must fit. To take an example from the National Library of Medicine MEDLARS System, each bibliographic reference is accompanied by between three and ten keywords, which are listed in no particular order with the citation to which they refer. These keywords are, however, selected from a controlled vocabulary of several thousand terms which form a complex structure of hierarchical levels, i.e. more specific and more general terms and also of synonym equivalence and cross referencing. In this case a large part of the structure which must be considered in order to make full use of the data is contained in a form which is entirely external to the data and, in fact, in the computer based MEDLARS System, the data itself is supplied and can be interpreted only in conjunction with a copy of the controlled vocabulary on magnetic tape in a form suitable for computer input. In the case of the American Chemical Society Data Base, the subject matter itself, i.e. chemistry, imposes a well-established structure, since the chemist interrogating the data base will know by long experience the relationships which exist between chemical compounds where the layman and the computer would not expect from observation of the names of the compounds that there was any connection. In many present day computer based systems, therefore, there is a structure which does not form part of the data base but which is nevertheless essential to enable full use to be made of the data base. In the same way that these aspects of structure are implicit in the data, it is commonly the case that the processing programs which require to take advantage of this structure also contain it implicitly within themselves. In other words, the programmers and designers who produced the computer based system had to have a knowledge of the implicit structure of the data and drew upon this knowledge when constructing the programs. It is very easy to lose sight of this fact when attempting to determine whether the processing programs can be of use with data bases to which the assumptions made about implicit structure in the processing programs may be completely inapplicable. The danger here is that, in transferring a new data base to existing processing programs, everything may, in fact, appear to go well and results may be obtained, but in practice the results may be biased by the implicit assumptions in the program, and may therefore not accurately represent the information which should be retrieved from the new data base. It is generally agreed that producing misleading information is a very much more serious fault than producing no information at all, and this aspect of data interchange must always be considered very carefully.

It will be seen, therefore, that any data item in a data base does not stand alone but must be considered in relationship to all the other items to which it is related. This relationship may be expressed in various ways; it may be implicit in the data and apparent only to the human readers, but in all cases to secure accurate and meaningful retrieval of information, these relationships must be considered, and if it is required to interchange data between information systems, then considerable benefit will be lost unless some means is found to transfer information about the implicit data structure in such a way that this can usefully be employed in the new system to which the data is despatched. Many systems employ retrieval criteria for selecting subsets of the data based purely on the original text stored within the computer and rely entirely on the user's knowledge of the interrelationships between terms to ensure that he obtains the information required. Such systems, of course, contain no processing element capable of recognising relationships between data items, and these systems, therefore, have a flexibility to handle a variety of input data bases at this very restricted level. As an example, the COSMIC Centre at the University of Georgia, which provides processing for a variety of data bases, can provide a selection service on ten or eleven data bases from different sources, as long as the necessary knowledge of structure implicit in the data bases is used in formulating the retrieval criteria.

It was mentioned in section 3.1 that the use of magnetic tape for data base storage imposes certain constraints upon the computer processing of the data. In particular, due to the serial nature of magnetic tape storage, it becomes impracticable to establish explicit links between related items, and such links must therefore be applied, either by the user in establishing the retrieval criterion which he wishes to use, or as part of the processing operation carried out on the retrieval criterion before applying it to the data base. As was shown in section 3.3, this constraint in fact facilitates interchange of data between different processing systems due to the enforcement of the separation between overall data base structure and the individual items comprising it. When we consider systems based on random access media (discs or magnetic card files), then we find that the distinction between the data base, its explicit and implicit structure and the constraints imposed upon data base processing by the physical properties of the computer become very much less clear. This is due to several factors; for instance, random access potential enables a much better use to be made of the overall data structure in establishing retrieval criteria, so that the overall data structure becomes, in fact, an implicit part of the data base and processing system. Secondly, random access storage is not infinitely extensible in the same way that magnetic tape storage is, and it is therefore necessary to consider very carefully the efficiency of utilisation of the storage medium in order to ensure that the system can handle a sufficient volume of data to provide the type of service necessary. Furthermore, many of the systems using random access storage do so because they offer a real time service, e.g. via computer terminals, and efficiency of data processing, from the computer point of view, is essential if undue delays in responding to a user's request are to be avoided. Where a data base has been established on the basis that capacity for providing explicit links between items is available, it is subsequently difficult to devise a method of representing the data which permits it to be placed on magnetic tape in serial form. In addition, the techniques adopted for storage and retrieval of data are normally so heavily oriented towards the processing facilities in use that it is difficult to transfer the techniques used other than to an identical computer. The data base and the processing system form an integrated whole which it is impossible to split into its constituent parts and which is so closely involved with the original computer that it cannot, without extensive alteration, be operated on any different computer.

4. DATA SELECTION

The overall structure of a data base, whether it be explicit or implicit, is normally closely related to the expected mode of use of the data. Any consideration of data structures must, therefore, include a discussion of the selection processes to be used to extract data for eventual use. Apart from those systems where the interest lies in presenting all the data in a variety of forms and formats, all other systems have some selection process by which a small subset of the data can be isolated as satisfying a certain combination of criteria. These may be general, such as the selection of special purpose bulletins from a broadly based data collection, or highly specific, such as the answering of enquiries from individual users of the system. Selection criteria normally take the form of specification of the elements which must appear in order that an item be selected, together with an expression of the relationship which must exist between such elements in the item. As an example, it may be necessary to distinguish between "the use of computers in education" and "education in the use of computers". Each of these examples would be indexed under the terms "computer" and "education" but, for retrieval purposes, it is necessary to know the relationship in which the terms stand in an item before determining whether it is relevant to the user's need. The recognition of relationships is very much more difficult than the simple recognition of existence or non-existence of terms, and in fact many existing systems do not offer the facility to make this kind of distinction. Techniques used to permit relationship expression range from simple juxtaposition or proximity criteria, for example, that specified terms shall be adjacent or shall lie within the same phrase, same sentence, or same paragraph; or each term when allocated to a document may have associated with it a statement of its function, i.e. as a primary descriptor in relation to the document, or as qualifying another descriptor. For retrieval purposes then the corresponding relationships can be expressed and matched. In our example given above, therefore, it may be possible to distinguish the two instances of the uses of the terms "computer" and "education" by indicating, both for retrieval purposes and when assigning terms to the document in the first place, which term is the main term and which term is a subordinate or qualifying term. Considerable work has been done in the United States of America on the possibility of processing text and by identifying the relationships between words as expressed in the normal grammar of the language to be able to develop relationships adequately for retrieval purposes. Unfortunately, the general conclusion seems to be that this technique loses its effectiveness just at the point where its discriminatory power would be most useful. In addition, the processing capacity required to be able to perform this analysis and corresponding matching is considerably greater than that required for other types of system, which increases the costs of such an information system by a large factor.

One interesting approach to the problem of retrieval criteria has been to make maximum use of such information contained with a journal article as has been supplied by the original author or the editor. Thus, many systems retrieve on the basis purely of information contained in the title, the author abstract, the summary provided by the journal editor, or by references to the article made in the citations of other articles. Evidence as to the validity of these approaches is conflicting, and the only firm conclusion which can be drawn is that the effectiveness of each of these methods of establishing retrieval criteria vary very considerably from scientific discipline to scientific discipline. As an example, titles alone seem to provide an effective mechanism for retrieval in the field of chemical literature. This may in part be due to the fact that a manual publication of chemical titles has been available for a large number of years, and there has therefore been an opportunity for authors writing papers in the field of chemistry to realise the value of concise, accurate and informative titles. In addition, the vocabulary of chemistry is controlled by a series of complex and widely accepted rules which again facilitate retrieval based purely on the vocabulary used by the authors of papers. The principle of citation indexing is that, by following citations to earlier papers, it is possible to identify significant papers by virtue of the number of citations made to them. In practice it appears that the very best papers and, in many cases, the very worst papers attract large numbers of citations which indicate that care must be used in relying on this method to identify significant papers in the scientific disciplines. As a general rule it seems apparent that the total amount of intellectual effort which must be applied to a document in order that it may be retrieved from an information system is constant; the variation occurs between the application of the intellectual effort initially when the data is included in the system through to systems where little or no intellectual effort is applied at input, but a correspondingly greater load is thrown upon the eventual user to define his retrieval criteria in such a way that relevant material will be selected. The economics of these approaches depend largely on the ratio between the volume of input and the volume of use. Any system where the volume of use tends to be large compared with the volume of data, obviously will be more economical if intellectual effort is applied at input stage. There is however the factor that it is easier to quantify and recognise expenditure at the data input stage than it is to quantify the cost to a user of less efficient retrieval or of an increased time spent in deducing appropriate retrieval criteria.

5. INFORMATION PROCESSING SYSTEMS

We have seen earlier that a considerable proportion of the investment in a computer based information processing system lies in the design and production of the programs which provide the processing facility. Apart from the cost of this exercise, there is an inevitable delay between a decision to set up a system and its ability to operate owing to the time necessary to implement the computer based system. There are therefore distinct economic advantages in an ability to exchange processing systems as well as data bases to enable a new processing centre to be set up, possibly for a new subject area, with a minimum of expenditure and delay. Where it is contemplated that an existing system should be used for a new application, two questions must be answered satisfactorily - first, does the proposed system provide the type of processing required in the new application? secondly, is it physically possible to operate the system on a computer available to the new processing system?

5.1. *Description and functions of a processing system*

To determine whether a potentially useful processing system will in fact meet the requirements, two things are necessary - a complete and accurate description of the function of the system under consideration, and a similar description for the system which it is required to provide. These descriptions are by no means as easy to come by as may be expected. A common failing of system descriptions is that they describe not the functions of the system but the use which is being made of it. This distinction does not become apparent until an attempt is made to assess the system's usefulness for application to a different data base. It then becomes obvious that the system description is geared to existing use and not to the actual capacities of the system. It is necessary to consider the form in which data is required for input to the system and to ascertain whether the data contents which are required in the new system can be successfully fitted into the data formats processed by the existing system. Next the validation and amplification processes need to be examined, as it is likely that validation in particular will be heavily application oriented and may be completely irrelevant for the new application. The selection and retrieval capabilities of the system must then be studied to determine whether these offer a range of facilities which would be adequate for the proposed use and how the needs of users of the new system could be met in terms of these facilities. Finally, forms of output, both in terms of presentation and medium, must be considered to determine whether they are acceptable to the expected users of the new system. There will be various constraints on the old system which may or may not be explicitly described, in particular constraints that are functions of the original system's use may not even have been recognised by the originators of the system, and there is a strong probability that their existence will be neglected or completely overlooked in the system description.

The second requirement for comparison is that the required functions and capabilities of the new system must in their turn be clearly defined together with any special requirements or constraints of input data vetting, storage, in particular with regard to volume and accessibility of data and the needs of system users. From this a description of the functions of the required systems can be deduced. Not until this stage has been reached is it possible to come to an informed decision as to the suitability of existing well described systems.

5.2. *Physical and economic constraints*

Having found one or more existing systems which appear to have the appropriate functional characteristics for the required processing system, we must consider various levels of physical compatibility. In the ideal case a computer identical with, or larger than, that used for the old system will be available for processing, and in this case it may be expected that with a minimum amount of trouble the old system may be operated on the available computer without change. This situation is, however, rare and, even so, the use of the system on a compatible computer with a different configuration may introduce inefficiencies which could be removed if the system were changed to take advantage of extra facilities available on the new machine, and also if necessary to make allowances for constraints imposed by deficiencies in the new machine. Such things as variations in the core storage available, number of peripheral devices, speed of tape units, availability of input and output devices, such as card readers, line printers etc. may make a system written for a machine from the same manufacturer inefficient on the configuration available. Indeed if the available configuration is appreciably smaller than that on which the existing system is being run, it may even not be possible to transer to the new machine. Factors to consider here, therefore, include the requirements of the system in terms of hardware, processor, storage available, number and type of peripheral devices, and also the interaction of the system with the operating regime of the machine.

It is more likely, however, that the machine available for processing the new system is not directly compatible with that used as the basis for providing the old system. In practice, this means that programs written for the existing machine cannot be transferred directly to the new machine and be expected to run since, with certain exceptions, the instruction set of a computer is unique to that model. This particular aspect of the problem of interchangeability is common to all forms of computer processing and has been apparent since the early days of computing. Some considerable effort has therefore been expended in an attempt to devise methods of transferring programs from one machine to another which does not have a compatible instruction set other than by a complete rewriting of the program. It would normally be possible to use the logic of the existing program, so that the cost of reprogramming for a new machine would be of the order of half the total cost of producing the original system. The principle is that it is possible to produce a programming language which is a formal description of computational processes in such a way that it is not closely related to the instruction set of any one computer. By virtue of the carefully defined nature of the language, it is possible for the computer itself under control of a suitable program to read instructions written in a high level language and to translate these into a form suitable for computer implementation. Thus, effectively, reprogramming is carried out when the high level program is transferred to another machine, but is carried out by the computer itself. It can commonly be expected that a fairly large program written in a high level programming language can be translated into machine instructions for a particular computer in a time not exceeding one hour, and commonly very much less, as compared with the three to six months it might take to rewrite the program by hand. These machine independent or high level programming languages tend to be developed to meet specific needs. Thus, there are languages (FORTRAN, ALGOL) which are specially developed to facilitate description of numerical processes and computations. Other languages (COBOL, CLEO) are intended for commercial data processing applications and are oriented towards the processing of files of coded and numeric information. Languages such as SIMULA and GPSS are intended to facilitate the simulation of numbers of independent parallel interacting processes such as transportation problems and machine shop scheduling. Considerable interest has been expressed recently in languages suitable for information processing, and the problem is being studied in various centres. There is, however, no generally acceptable language specially oriented towards this particular problem at the moment. The use of high level machine independent languages is not without its problems; for ease of interchangeability a language must be available on a considerable range of machines, and in addition it is necessary to ensure compatibility between the implementations of the language on these machines. Thus, although a program may be written in COBOL for one computer, it may require small but significant alterations to be run on a different computer with a translator for ostensibly the same language. With care, however, this problem can be overcome without much difficulty and without much expense. The second problem is that in introducing an intermediate language between the problem and the computer an extra potential source of inefficiency is introduced. Two major points at which inefficiency occurs relate to the basic suitability of the programming language for the problem which it is required to solve, and secondly the basic efficiency with which the machine concerned can execute programs written in the high level language chosen. It is commonly possible when writing in the machine's own instruction set to take advantage of special facilities on the machine to improve the efficiency of the processing program. There is as usual no simple solution; it is necessary in each case to compare the costs of writing a program in the machine language,which are generally higher than the costs of writing in a machine independent language, with the increased running costs to be expected due to the relative inefficiency of programs written in a machine independent language. Due weight must also be given to the requirement for interchangeability of programs across computers. A number of information processing systems have been written, and operate successfully and economically, in the languages FORTRAN and COBOL. In many instances it is possible to take advantage of a facility whereby parts of the program that can contribute significantly to the efficiency of the system are written in the most efficient way possible using machine code, whereas the rest of the system is written in a machine independent language. Translation of such a system to a different computer will then involve producing for the new computer in its own instruction set replacements for those parts of the system written in the machine language of the old computer, and as long as these insertions perform the same function on the new computer, they will interface successfully with the parts of the system written in machine independent language. The success of this technique depends heavily on accurate and thoughtful documentation of the functions to be performed by those parts of the system written in machine instructions, and an accurate description of the way that these parts interface with the rest of the system. The efficiency to be expected from a system written in machine independent language is normally between 50-75% of that of the system written in machine language, depending largely on the extent to which input and output are the limiting factors in processor time. This increase in the running cost of the system must be weighed against the possible advantages in terms of saving in development costs and the cost of delay in setting up the system in each individual case. The use of an existing system written in high level language can save between two and ten man-years of programming effort and the corresponding delay involved.

It has been mentioned above that, to make the most efficient use of systems based on random access storage, it is necessary to base the system design very closely on the characteristics of the storage device concerned. In this respect the use of machine independent languages can introduce a much more serious loss of efficiency, as the facilities provided for accessing random access devices cannot take advantage of special characteristics of the data base or of the processing required to be performed on it. Generally speaking, the more advantage a system takes of the particular facilities available on the machine for which it was first produced, the less likely is it that it can be transferred, or transferred efficiently, to a different computer.

6. CONCLUSION

In summary therefore the following points must be considered when investigating the potential for interchange of data or systems:-

1. *Physical compatibility of storage media for data interchange*

   In practice, the most likely way of achieving this is by the use of generally accepted standards for magnetic tape.

2. *Logical compatibility of explicit data structures*

   Here, growing interest is being expressed in adopting something like the USASI standard for information interchange as an international standard to facilitate exchange of data in a form suitable for supplying information services.

3. *Compatibility of, and processing capability to accept or ignore, implicit data structure*

   Careful consideration must be given to whether an implicit data structure can efficiently be ignored without seriously affecting the potential of the data base for retrieval.

4. For systems interchange it is essential that we start with a clear understanding of the functions which it is required that a system should perform. It is then possible to consider the compatibility of functions provided by potential systems.

5. At a lower level we must consider program compatibility, considering whether a potential program will actually run on computing facilities available, or, if it is written in a suitable high level language, whether the efficiency of using the language is acceptable.

6. *Compatibility of configuration*

   A system which requires a very large extensive and expensive computer on which to operate is of little use if such a configuration is not available elsewhere, even if a computer of the same model is accessible.

It will become apparent that, when seeking to exchange data or computer based systems, the most important considerations relate to the potential of the existing or donor system for compatibility, as a lack of interchangeability in the source can seldom be corrected at the destination. It is generally held that for systems or data to be interchangeable it is necessary for the constraints and requirements of interchangeability to be borne in mind right from the initial stages of development of the system. Since the constraints of interchangeability will in general increase the implementation cost and possibly reduce the running efficiency by a small amount, there is obviously a strong economic incentive for systems to be produced in a way which is inimical to interchangeability unless the potential advantages of subsequent interchangeability are borne in mind and the requirements of interchangeability are imposed as part of the original design requirement.

Information processing by computer is currently suffering from the parochial nature of most systems, which were designed before considerations of data interchange or systems interchange had been considered important. As a result of this, it will obviously be necessary to generate many new information processing systems entirely independently. It is to be hoped that the benefits to posterity of incorporating the potential for interchange in these systems will be sufficiently widely known as to make it unnecessary to repeat this process for the third generation of information processing systems.

APPENDIX

International and other Standards relevant to compatibility

ISO 646-1967	6 and 7 bit coded character sets for information processing interchange.
ISO 961-1969	Implementation of the 6 and 7 bit coded character sets on 7-track 12.7mm (½ inch) magnetic tape.
ISO 962-1969	Implementation of the 6 and 7 bit coded character sets on 9-track 12.7mm (½ inch) magnetic tape.
ISO 1001-1969	Magnetic tape labelling and file structure for information interchange.
ISO 1418 draft	Representation of six and seven bit coded character sets on punched tape.
ISO 1538 draft	Programming language ALGOL.
ISO 1539 draft	Programming language FORTRAN.
ISO 1671 draft	Dimensions for punched paper tape for data interchange.
ISO 1299 draft	Flow chart symbols.
ISO 1679 draft	Representation of ISO 7 bit coded character set on 12-row punched cards.
ISO 1681 draft	Specifications for unpunched paper cards.
ISO 1682 draft	Dimensions and location of rectangular punched holes in 80-column punched paper cards.

ISO 1729 draft	Properties of unpunched paper tape.
BSI 68/11624	Part 4 - General requirements for the interchange of data on punched paper tape.
USASI Z.39.2	Draft USA Standard for a format for Bibliographic Information Interchange on magnetic tape.

BIBLIOGRAPHY

BEMER, Robert W.
Program transferability
in Proceedings of the AFIPS 1969 Spring Joint Computer Conference
pp.606-607.
AFIPS Press, Montvale, N.J., U.S.A.

BRODERICK, William Russell, and BARKER, P. J.
Basic Algol.
Iliffe Technical Publications Ltd., London, England.

DAHL, Ole-Johan and NYGAARD, Kristen.
Basic concepts of SIMULA, an Algol-based simulation language
in Digital simulation in operational research.

GOSDEN, John A.
Software compatibility
in Proceedings of the AFIPS 1969 Spring Joint Computer Conference,
pp.607-608.
AFIPS Press, Montvale, N.J., U.S.A.

LYSEGARD, Anna.
Introduction to COBOL.
Oxford University Press, London and New York.

MORENOFF, Edward.
The transferability of computer programs and the data on which they operate
in Proceedings of the AFIPS 1969 Spring Joint Computer Conference
pp.609-610.
AFIPS Press, Montvale, N.J., U.S.A.

MURRIL, Paul W. and SMITH, Cecil L.
FORTRAN IV Programming.
International Textbook Company, Scranton, Pa., U.S.A.

PROBLEMS OF THESAURUS CONSTRUCTION FOR EDUCATION

by

H H STRÄTER

Preliminary note 25

Educational terminology 25

Thesaurus problems illustrated by means of an example 26

Systematic classification in a thesaurus 33

The *Information Retrieval Thesaurus of Education Terms* 36

Guidelines for constructing a thesaurus 40

Pre-requisites for the construction of a thesaurus 41

Appendix 1. *Thesaurus of ERIC Descriptors* - translation from English into German 45

Appendix 2. Comparison between the General Headings of one facet of the IRT and their equivalents in a projected German Thesaurus 47

Bibliography 49

1. PRELIMINARY NOTE

This study is based upon research work originally carried out with the aim of investigating the possibilities for constructing a *German* thesaurus for the field of education. Certain specific prerequisites (such as the terminology available) do not lose their significance even when considered from the aspect of European co-ordination as they enable linguistic comparisons to be drawn between various languages (cf. Appendix 1). In this study the view is put forward that it is important not only to elucidate the theoretical bases, but also to test them in terms of projects already carried through.

It will only be possible to start from a very limited area of thesaurus development, the comparison and evaluation of two English language thesauri developed for education with a German authority list that was already available. The English language thesauri in question are the *Thesaurus of ERIC Descriptors: Phase I* (June 1966)(second edition April 1969) (1) and the *Information Retrieval Thesaurus of Education Terms*, Cleveland (1968). The German authority list has been in use since 1966 for the *Bibliographie Pädagogik* (2).

This study does not intend to present an evaluation of all available classification systems in education and therefore omits, for instance, the London Education Classification as well as the French classification system and others.

The study treats the following questions: it starts with the problem of educational terminology, 2, illustrates various thesaurus problems by means of the *ERIC Thesaurus*, 3, it then tries to present the problems of a systematic classification in a thesaurus, 4, introduces the *Information Retrieval Thesaurus of Education Terms*, 5, and gives a few guidelines, 6, as well as prerequisites, 7, for thesaurus construction.

The following three questions are also dealt with in chapters 3 - 6:

- To what extent is it possible to translate a thesaurus system from one language into another (using English and German as a basis for comparison)?
- What are the chief results yielded by thesaurus research so far?
- Are mechanical systems already in existence with which it is possible to produce a thesaurus automatically?

The answers to these questions can by no means be regarded as complete; rather, in the course of their examination, new problems have become apparent: a main contributory factor is the uncertainty of conditions in the educational field (which is developing in the direction of the social sciences) as well as in the area of thesaurus research.

2. EDUCATIONAL TERMINOLOGY

As well as theoretical discussions of thesaurus problems, reports based on experience gained in practical work can be found in the technical literature. There is really no lack of these reports, but those that are available are almost exclusively devoted to thesauri (or unfinished thesauri) which use a technical discipline or a natural science as a basis. However the relationships between concepts in the humanities are much more complex. The demarcation of a term in its field is, as a rule, considerably more difficult than in the natural sciences or technology. The authors of the *ERIC Thesaurus* designate educational terminology, appropriately, a 'soft' language (3). Nevertheless, the resignation occasionally found in the face of difficulties encountered in classifying and systematising terminologies in the humanities would appear to be out of place. It should be stated first of all that we find ourselves at the moment just at the beginning of a development, the possibilities of which can hardly be exhausted in logical abstractions, but which rather come to fruition in practical experimentation, well thought out and controlled. Moreover, Roget, Wehrle/Eggers and Dornseiff (4), in creating concept systems for the English and German languages, have shown that it is possible to determine the position of a concept in its context in the humanities. (Despite the many shortcomings of these thesauri, they have proved their practicability.) Analysis of these thesauri might, therefore, promise information also relevant to the education field. These thesauri are not planned from the point of view of information retrieval, however they nevertheless suggest that, with regard to systematic classification, that which is possible for the entire area of a language must also be possible for an area more tightly defined, such as that of education.

---

(1) See 3.2.

(2) See 3.3.

The analysis of the relationships between concepts in the educational field is work which has yet to be done. The following pages are concerned essentially with experimentation in the English language. At the same time the aim is to evaluate the various initial work done as objectively as possible, without being tied down right at the outset to a set of classifying principles.

3. THESAURUS PROBLEMS, ILLUSTRATED BY MEANS OF AN EXAMPLE: THE *THESAURUS OF ERIC DESCRIPTORS* (5)

3.0 *Introduction*

Classification systems in the form of thesauri are being developed at the moment in the most varied technical fields. A detailed survey of these projects is however not possible here.

As a basis for the present study, eleven thesauri were analysed. The decisive factors in the selection were, first, the diversity of the design and organisation of the thesaurus as it was necessary to compare varying models, and secondly, the need to relate to the educational field.

The purpose of this investigation was, essentially, to test the possibilities, under the given conditions, for creating a thesaurus for the field of education. For this reason a detailed analysis of the thesauri was dispensed with; rather, they have been collated in a summary according to primary characteristics (see Fig. 1). The *ERIC Thesaurus* and the *Information Retrieval Thesaurus of Education Terms* (IRT) were subjected to detailed analysis as they both concern the field of education. So that the prerequisites for thesaurus construction could be examined in greater detail, a section discussing the basic problems involved in classification systems has not been included. Only a few references are made here to the constitution of the thesaurus concept.

---

(3) *ERIC Thesaurus: Phase I*, p. XIII.

(4) See Fig. 1.

(5) *EDITOR'S NOTE:* This section contains an analysis of the *Thesaurus of ERIC Descriptors* based, as the author states, on the preliminary *Phase I* version (June 1966). Since *Phase I*, an *Interim Edition* (January 1967), a *First Edition* (December 1967) and a *Second Edition* (April 1969) have been released, and therefore certain criticisms made or implied here may no longer be valid. We feel justified, however, in printing the section as it stands as being more illustrative of the problems likely to be encountered by the thesaurus constructor than it would be if it were based on more 'perfect' material.

Derived from the Greek *thesauros* (treasure, storehouse) the term 'thesaurus' had its origins in the Middle Ages and was originally a quasi-synonym for 'dictionary' (6). The first thesaurus in the sense of the word as used today is that by P. M. Roget, *Thesaurus of English Words and Phrases*, the first edition of which appeared in 1852. The *Chemical Engineering Thesaurus* (American Institute of Chemical Engineers, 1961) is regarded as the first information retrieval type thesaurus. Although Roget thus provided the foundation upon which all the present thesauri are based, the aspect of information retrieval has been added and made new forms of organisation necessary.

By 'thesaurus' shall therefore be understood any word list in which the words are classified according to interterm relationships. Such a thesaurus may fulfil the following functions:

- It forms the basis for fixing subjects,
- It acts as a control for the classification system (this function is to be understood, of course, also in the reverse sense),
- It acts as a basis for terminological investigations (frequency of use of certain terms, etc),
- It does valuable work for scientists by virtue of the fact that it not only contains related words but it also yields information about the meanings of words and about systematic relations in a particular area.

3.1 *The course of the investigation*

Two questions were to be clarified in examining the *ERIC Thesaurus*:

- Is it possible to translate the *ERIC Thesaurus* into another language (the basis for comparison was German)? It is obvious that, if this were possible, a great saving in time and money could be made. (It is conceivable that a single translator, if able to consult experts in the field occasionally, could manage the work); this would be a significant advantage over establishing a thesaurus afresh in another language.
- How high is the standard of the *ERIC Thesaurus*? (compared with other thesauri available; is it practicable etc?).

This section is mainly devoted to the first question. The answer to the second must, in many respects, remain hypothetical, but in the final part of the section an attempt is made, from the angle of the structure of the *ERIC Thesaurus*, to arrive at a tentative estimate of its standard.

The investigation of the translatability of the *ERIC Thesaurus* involves the examination of the terminological problems, and of the relationships between the terms and their positions with regard to one another, that is, the systematic problems.

3.2 *Details of the ERIC Thesaurus*

The first version of the *ERIC Thesaurus* appeared in 1966 (*Thesaurus of ERIC Descriptors: Phase I*, June 1966, US Department of Health, Education and Welfare - for review purposes only).

This study is based mainly on Phase I as the second edition was not received until the study was nearly completed. The second edition of the *ERIC Thesaurus*

---

(6) For the etymology of the word 'thesaurus' and the historical development of the term 'thesaurus', see MODEL F: *Thesaurus in der Dokumentation* (see bibliography).

Fig. 1: A summary of various thesauri according to their main characteristics

Title	Subject	Arrangement	Reference Words	Editor
Roget's Thesaurus of English Words and Phrases, 1852 (first edition)	English language	Two sections: (1) System of classification (= 6 main classes) (2) Alphabetical Index	see	P. M. Roget
Deutscher Wortschatz, 1866 (first edition)	German language	Two sections: (1) System of classification (= 6 main classes) (2) Alphabetical Index	---	Schlesinger, Wehrle, Eggers
Thesaurus of ASTIA Descriptors, 1960 (first edition)	Military affairs	Three sections: (1) Schedule of Descriptors (26 "Descriptor fields") (2) Scope Note Index (3) Charts of Generic Relationships	also see specific to generic to includes	Armed Services Technical Information Agency, Arlington
Descriptor List 1963 (first edition)	Atomic energy	Three sections: (1) Materials List (2) Subject List (3) Reactor Names, Designations, and Types List	see see also	Atomenergie-Dokumentation beim Gmelin-Institut, Frankfurt
Der deutsche Wortschatz nach Sachgruppen, 1933 (first edition)	German language	Two sections: (1) System of classification (= 20 main classes) (2) Alphabetical Index	---	Franz Dornseiff
Thesaurus of Engineering Terms, 1964 (first edition)	Engineering terms	Alphabetical Index	use used for narrower term broader term related term	Engineers Joint Council, New York

TDCK - Circular Thesaurus System, 1964 (third edition)	Military affairs	Two sections: (1) Circle-scheme (2) Alphabetical Index	---	Netherlands Armed Forces Technical Documentation and Information Centre, Den Haag
EURATOM-Thesaurus, 1964 (first edition)	Nuclear energy	Four sections: (1) Alphabetical lists (2) Glossary (= more than 2,000 non-keyword terms) (3) Keyword groups (4) Graphic display schemes	use use...+ (=and) see see...or	EURATOM
Thesaurus of ERIC-Descriptors: Phase I, 1966 (Second edition 1969)	Education	Alphabetical Index	use used for narrower term broader term related term	US Department of Health, Education, and Welfare
Programmietter Unterricht Thesaurus, 1966 (first edition)	Programmed instruction	Alphabetical Index	siehe nicht zu verwechseln mit nicht benutzter Begriff Oberbegriff Unterbegriff verwandter Begriff	Pädagogisches Zentrum, Berlin
Information Retrieval Thesaurus of Education Terms, 1968 (first edition)	Education	Alphabetical Array Faceted Array Permuted List of Descriptors	use used for broader term narrower term related term (+ variant) scope note (+ variants)	Barhydt and Schmidt with Chang

contains 6,251 descriptors, including all synonym entries. Although it is an improvement on Phase I, there is little change in basic structure. Nevertheless it is most interesting that the second edition contains a 'Descriptor Group Display': a first step in the direction of a taxonomy.

For the first version a working period of nine months was needed. The authors regard this first version as a "first step in the development of a structural compilation of educational terms" (7).

The terminology incorporated in the thesaurus was compiled using 1,740 documents; these covered only a segment of the educational field: ". . . the 1,740 documents from which this thesaurus was developed concern the education of disadvantaged children mainly at the pre-school, elementary and secondary level" (8).

We are here concerned with an experiment, with a first approach to the problem of constructing a thesaurus; on the other hand it is definitely a very extensive experiment for, as already pointed out, the network within the educational terminology is very great and, consequently, we can regard these terms to a high degree as being representative.

Fundamentally, the authors acknowledge their adherence to the method of 'co-ordinate indexing'.

The thesaurus was constructed in three stages:

(i) Processing the documents according to the 'free indexing' approach; thus the work was carried out without a systematic list of terms (authority list).

(ii) Alphabetical classification of the terminology thus listed.

(iii) Testing of the terminology by groups of specialists (educational experts as well as documentation experts) and establishing relationships between the terms ('structuring the terminology') (9).

The method of free indexing was used throughout except when certain 'broader terms' were formed (for reasons of logical connection).

The model used by the authors was the *Thesaurus of Engineering Terms*.

The number of descriptors in the Phase I thesaurus amounted to 2,379; synonyms and quasi-synonyms totalling 881.

3.3 *Terminological problems (translation problems)*

The author of this study translated approximately 20 per cent of the terms in the *ERIC Thesaurus* into German, as follows:

(a) All the terms listed under the letter A, so as to be able to compare the exact position of the descriptors in the thesaurus (together with their reference words) with the conditions in Germany.

(b) Approximately one hundred 'broader terms' from the aspect of the comparability of American and German broader terms;

(c) Approximately fifty broader terms with their 'concept-ramifications' in the thesaurus; but, however, by no means all their 'concept-ramifications', for that would have involved a translation of the entire thesaurus, as will later become evident. With regard to this point the comparability of the concept network in German and American was to be tested.

---

(7) *ERIC Thesaurus: Phase I*, p. III.

(8) Ibid, p. V.

(9) Ibid, P. VIII.

The selection of broader terms was made with the aim of obtaining, as far as possible, extensive 'far-reaching' broader terms, but otherwise without restriction. The translation was done without consulting any specialists and the terms thus translated were compared with a list of words proposed by the *Dokumentationsring Pädagogik* (10); the terms listed under the letter A were compared word for word, the broader terms as seemed desirable.

Appendix 1 contains translation examples together with words (for comparison purposes) from the list proposed by the *Dokumentationsring Pädagogik*. At this stage we can content ourselves with a summary of the results of the terminological examination:

(i) The translation should be carried out by someone possessing a basic knowledge of educational terminology, especially with the assistance of a subject index similar to that of the *Dokumentationsring Pädagogik*, and acting in collaboration and consultation with specialists in the field (11).

(ii) Terms which are relevant only to American education (e.g. negro colleges) should be eliminated promptly. However eliminating all the concept references in the thesaurus involves considerable trouble and a noteworthy loss of time, it is, moreover, a process with many hidden sources of error. This fact emanates primarily from the structure of the *ERIC Thesaurus*, which will be discussed in more detail later; so that there are as many entry-points as possible when referring to the thesaurus, its reference system is very comprehensive.

(iii) The greatest difficulty however is that of fitting into the thesaurus terms relevant to German education but which are not present in the *ERIC Thesaurus*. Although this is not a problem primarily connected with the translating issue, it crops up as an immediate consequence.

Without however the aid of lists of terms in French, etc. constructed by specialists, it is hardly likely that one could manage. Then, in addition, the following question is relevant: would it perhaps be more advantageous or efficacious to construct thesauri from scratch in German, French, etc? Of course the *ERIC Thesaurus* could then serve as a valuable aid; as to whether it can be regarded as a yardstick is a question which has still to be clarified and which will be dealt with in the next section.

3.4 *The structure of the ERIC Thesaurus*

The structural problems of the *ERIC Thesaurus* in their positive and negative aspects are explained here by means of an example from the thesaurus (Fig. 2). Starting with a narrowly-defined term taken at random from the thesaurus ('classroom furniture'), the ramifications of this term in the thesaurus are made apparent; that is to say, what concepts are arrived at, having started from this one term. In this particular case, we are dealing with a compound term. However other experiments

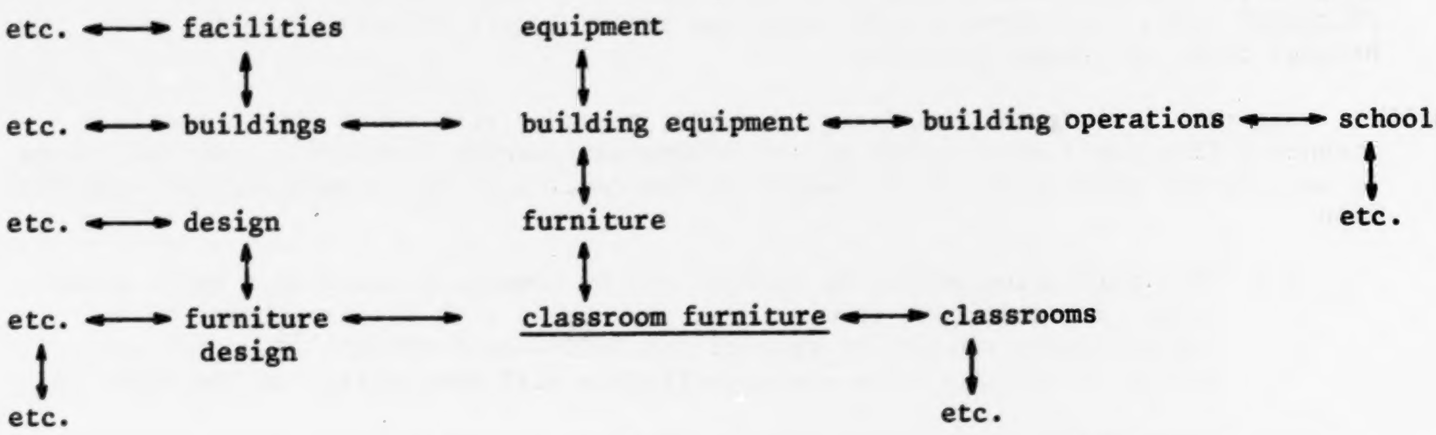
---

(10) *Dokumentationsring Pädagogik* is a chain of institutions in the Federal Republic of Germany documenting in the field of education.

(11) According to statistics of the Consultant Bureau, New York, a trained human translator can manage a daily output of approximately 2,000 words - quotation from PIETSCH E H: *Die zukünftige Entwicklung in der Dokumentation*, Libri 12 (1963) 4, p. 298. For translating thesaurus terminology however, a much lower daily output must be set.

Fig. 2: An example to illustrate the structure of the **Thesaurus of ERIC Descriptors**

"classroom furniture":
broader terms, narrower terms and related terms



↔ = related terms

↕ = narrower term/broader term

From the *Thesaurus of ERIC Descriptors: Phase I,* the second edition gives as related terms: classrooms, furniture arrangement, furniture design.

which were carried out with simple terms inevitably yielded the same results as the example given here. Almost every word exhibits an ambiguity of meaning (just as, on the other hand, every word has a primary meaning which makes it possible to give it a classifying position in a system).

In our case therefore the initial point is the narrowly defined term which then leads to the broad concept, as depicted in Fig. 2. At all events from the concept 'schools', with all its references (which cannot be presented here), the paths take us through the entire reservoir of the thesaurus. However we are also led into vast areas of the thesaurus via 'buildings', and its references. Evidently the conscious aim of the authors is to have as many 'entry-points' to the thesaurus as possible.

The many ramifications originate on the one hand as a result of hierarchical elements (broader terms, narrower terms), but also due to lateral ramifications (related terms). The problems of the latter term are the consequence of its lack of clarity; moreover, as the authors deliberately dispense with authority lists and, as indicated, only insert hierarchical elements here and there, the logical classification must appear, to a certain extent, questionable. Below is listed a case from our example:

'furniture design' is a related term to 'furniture'

'furniture design' is a related term to 'classroom furniture'

It therefore follows that 'classroom furniture' must be a related term to 'furniture'; however in the thesaurus 'furniture' and 'classroom furniture' are listed as being reversible and having a hierarchical relationship ('broader' - 'narrower term'). This incorrect relating of terms is not an isolated case and certainly not just a mistake which can be ascribed to the experimental nature of this stage of the thesaurus project, but rather it is inherent in this system.

In the *ERIC Thesaurus* the difficult problem of a taxonomy has been avoided. By using related terms extensively, the need to fix the relations between terms has been by-passed. The indefinite meaning of this term makes it possible to depict the most varied relations. It is a 'use' reference just as much as a reference word for hierarchical relations.

An argument used in favour of this method of procedure is that it is only important to have as many entry-points as possible to the thesaurus. However as there is no systematic general summary, many relevant terms cannot be found or are found only after time-consuming searches involving numerous sources of error, which cannot be afforded in daily work.

Apparently the authors of the *ERIC Thesaurus* are of the opinion that a taxonomy, for education at any rate, is not at all possible; first of all, as already mentioned, this is inherent in the very terminology of education: "Some of the problems encountered reflect directly upon the problems of educational language, which is, relatively speaking, a 'soft' language. By contrast, the subject matter of the sciences and engineering is more refined, lending itself to an accepted interpretation of language and word usage to an extent which is not apparently existent in the social sciences" (12).

Secondly, the authors find it virtually impossible to clarify the relationships between terms in the educational field; it is for this reason that so many compound terms were incorporated in the thesaurus: "This can be explained in terms of the educational process itself. Basically, this process involves students, teachers, and an educational environment; and the events which occur during this process can often be equally applicable to either the students, the teachers, or the environment" (13).

This statement, undoubtedly correct, should not induce us to deny fully the possibility of separating the terms and analysing them lucidly. It is true that a more distinct separation of the terms can be established in the natural sciences, but here too the most varied aspects of a matter, process or object often come together. Thus in an agricultural thesaurus, 'beetroot' should be regarded from the viewpoint of the farmer, the food chemist, the consumer, the shop-keeper; and these aspects can occur together and overlap.

All authors of technical and scientific thesauri have decided in favour of dividing the thesaurus into at least two sections, an alphabetically arranged index and a systematic general summary.

Thus certain drawbacks in the structure of the *ERIC Thesaurus* can be noted; these could be avoided by a well thought-out taxonomy.

4. SYSTEMATIC CLASSIFICATION IN A THESAURUS

4.1 *Assumption*

This is not the place to discuss the far-reaching problems of various systems; rather, by means of projects already carried out it is intended to measure certain standards required of a taxonomy. "Every taxonomy represents a classification which requires a considerable effort in drawing it up. For reasons of economy the work involved in constructing such a taxonomy should not go beyond the requirements demanded by the user. Classification systems are an aid to work, but not an end in themselves" (14).

---

(12) *ERIC Thesaurus: Phase I*, p. XIII-XIV.

(13) Ibid, p. XIV

(14) HUNGERMANN E H: *Zur Planung, Erstellung und Bewertung von Dokumentationssystemen*, p. 153 (see bibliography).

Delimiting the user requirements, of course, involves considerable difficulty, especially in our case.

4.2 *Organising a taxonomy in a thesaurus*

"Actually no one is truly comfortable in the role of a Decalogue-maker, not even Moses" (15).

It is an accepted fact that a purely mono-hierarchical structure cannot be established for either the humanities or the natural sciences. It is just as clear on the other hand that "a hierarchical minimum is essential for the formation of groups of concepts" (16).

Thirdly, it would appear just as necessary that the thesaurus contain, as well as the alphabetically-arranged index, a further control instrument.

The taxonomy guarantees control over the alphabetical index and thus, with a high degree of certainty, prevents unreferenced synonyms or near-synonyms being incorporated into the alphabetical index during up-dating. Useful preliminary work for EDP is thus realised, as certain sections of the field of work are demarcated. This is important for searches which can be limited (17).

How this control instrument should be constructed is another question. In this respect various experiments have been attempted, however reports dealing with their practicability have seldom been published. If we turn to the educational field, it would appear necessary to have a taxonomy that does justice as completely as possible to the complexity of conceptual relationships in educational terminology. This demand is easily made but very difficult to meet. As yet, the only thing that has become quite clear is that 'classical' classification hardly gets us anywhere and rather, that an autonomous classification is necessary for the documentation.

In addition, it is necessary to fix the relationships between the descriptors precisely, as these relationships are not inherent but must first be brought into being (18).

In this connection reference is made to a system which makes use of a faceted classification in order to obtain a fixation of relationships: D. J. Foskett's *London Education Classification* in *Education Libraries Bulletin*, Supplement 6, London 1963.

"However it is not intended to deal with this system at further length here; its practicability for computer-based information retrieval has yet to be tested. Fig. 3 shows a comparison between one position of Decimal Classification (Dewey) and Foskett's Faceted Classification (19)."

In section 5 we shall be having a look at another attempt to construct a taxonomy for a thesaurus.

---

(15) TAUBE M: *Extensive Relations* . . . (see bibliography).

(16) SCHEELE M: *Ein Verfahren zur automatischen Klassifizierung* . . . (see bibliography).

(17) Cf. BAUER G: *Die Bedeutung der Kategorien* . . , p. 62 (see bibliography).

(18) Cf. DOBROWOLSKI Z: *Analyse der Klassifikationssysteme*, and BLANKENSTEIN G: *Die hierarchische Struktur eines Thesaurus* (see bibliography).

(19) To make the reservoir of descriptors clear, graphical representations have also been used occasionally, as, for example, in the TDCK - *Circular Thesaurus of the Netherlands Armed Forces Technical Documentation and Information Centre* (Concentric circles).

Fig. 3: *London Education Classification and UDC* (Dewey)
e.g. "nature study in the kindergarten"

London Education Classification		UDC	
Ral	Schools and schoolchildren general	372.2	Elementary School Organisation
Ram	Nursery and infants, 3-7 years	372.218	Kindergarten
Ran	Nursery	372.3	Science and nature study
Rar	Kindergarten	372.4	Reading and spelling
Mab	Curriculum general	372.7	Arithmetic
Mim	Mathematics		
Mob	Sciences		
Moj	Biological sciences, "nature study"		

Nature study in the kindergarten = Rar Moj
Mathematics in the kindergarten = Rar Mim

Nature study in the kindergarten = ?

From FOSKETT D J: "How to find out educational research", p. 43-44.

4.3 *Digression: Thesauri and Electronic Data Processing (EDP)*

Electronic data processing and information retrieval by computer have made clear the necessity for a thesaurus for the purpose of documentation in a given field of work and have yielded new information for the construction of thesauri. This section can be regarded merely as a digression since extensive treatment of this theme would require a complete, separate report.

Without doubt it is basic to keep in mind the requirements of a future thesaurus constructed by means of EDP units, or at least the requirements for a revision of a thesaurus (20).

Experimenting with EDP units has clarified even further a fact which had been realised earlier - a taxonomic classification is absolutely necessary as a control instrument for a thesaurus. Automatic processing of texts has yielded new information - for non-automatic processing as well (21). However the construction of a thesaurus by automatic means is not possible at present. IBM has presented (for its system 360) a program for constructing a 'thesaurus' (22). The reservoir of words (thesaurus) fed in is not tied down to this system for ever.

By storing the non-numerical concepts from the texts internally, it is always possible for the 'thesaurus' to incorporate new concepts, and information already stored can be searched through (according to the concepts). Synonyms can be easily characterised (homonyms however pose extreme programming difficulties). Logical connections by means of 'and', 'or' and 'not' are possible. Nevertheless, this does not make a thesaurus. The relationships between words (with regard to meaning) must be fixed beforehand and even when being re-incorporated cannot be recognised by the machine. Only quantitative criteria are taken into account. On the other hand semantic qualities cannot be registered by the machine.

---

(20) See COATES E J, elsewhere in this Volume.

(21) For example, see LUHN H P: *Auto-encoding of Documents* . . .

(22) *ANIS* (see bibliography) and DÜRR H G: *ANIS* (see bibliography).

All the same new possibilities are becoming apparent here; for example, Siegfried Wagner, in his dissertation *Automatische Stichwortanalyse nach dem Rangkriterionverfahren* (23), has put forward a method by which the computer, with the aid of certain criteria fed in, selects words not only according to their frequency but also to their 'content' - that is, their semantic quality. It is not known to what extent such programming work is feasible at the moment in industrial production.

5. THE *INFORMATION RETRIEVAL THESAURUS OF EDUCATION TERMS* (IRT) (24)

5.1 *The development of the IRT* (25)

Work at the Center for Documentation and Communication Research was begun in 1961. "By early 1965, when actual work on the education thesaurus project was begun, sufficient knowledge and experience had been accumulated in these two areas (an analysis of document and question languages and a survey of the types and forms of information which users want) to provide a firm foundation for the preparation of a thesaurus that would serve the needs of an information system for education."

From 1961-1965 more than 7,500 documents were indexed. "Index terms assigned to these documents were collected in a coded word list or *Semantic Code Dictionary of Education*, and it was from this dictionary that the initial selection of terms was made." The *Semantic Code Dictionary* contained about 10,800 terms. Some classes of terms were excluded. Among these classes "are the names of all specific tests (except statistical tests) and the names of specific institutions and corporations, theories, programs and legislation. Geographical terms were considered to be identifiers and, with some exceptions (e.g. rural areas), were excluded from the completed thesaurus." At the end about 5,000 terms remained. "Because of the emphasis on newer media research in the document collection, certain areas of terminology were not very well represented in the working list. An attempt was made to fill those gaps so that the thesaurus would be reasonably representative of all areas of education."

5.2 *The structure of the IRT*

The *Information Retrieval Thesaurus of Education Terms* is in three parts: an alphabetical array, a facted array, and a permuted list of descriptors. As an example, the thesaurus entries for 'ability' (the second term in the alphabetical array) are represented in Fig. 4 as they appear in each of the three parts. The asterisked scope note (SN *) identifies terms which, in general, lack specificity and usually are dependent on context to establish their meanings. It would be better to use a more specific term.

5.3 *Testing the IRT (Problems of polyglot thesauri)*

5.31 *The field of programmed instruction*

We reduced a first test of the IRT faceted array to the field of programmed instruction. A German thesaurus of programmed instruction existed (26). Thus, the approach was from the German programmed instruction vocabulary. We did not translate the IRT (27).

---

(23) See bibliography.

(24) "Information Retrieval Thesaurus of Education Terms", ed. by BARHYDT G C and SCHMIDT C T with the assistance of CHANG K T, Cleveland: The Press of Case Western Reserve University (1968).

(25) Ibid, p. 2-3.

(26) *Programmierter Unterricht. Thesaurus.* Berlin: Pädagogisches Zentrum (1966). (No taxonomy; alphabetical index only.)

(27) See 3.3.

Fig. 4: An example to show the structure of the Information Retrieval Thesaurus of Education Terms

Alphabetical Array
IRT Part 1

```
ABILITY
     SN  *
     RT    6001
         ABILITY GROUPING@
         ABILITY TEST@
         APTITUDE@
         HANDICAP
         INDIVIDUAL DIFFERENCES@
         SKILL@
```

Faceted Array
IRT Part 2

```
      FACET          6
      SUB-FACET   6001

ABILITY
AVERAGE INTELLIGENCE
EDUCABILITY
GENERAL ABILITY
GIFTEDNESS
HIGH INTELLIGENCE
INTELLIGENCE
INTELLIGENCE QUOTIENT
IQ
LOW INTELLIGENCE
MENTAL ABILITY
MENTAL AGE

ILLITERACY
LISTENING ABILITY
LITERACY
MECHANICAL ABILITY
MUSICAL ABILITY
NUMERICAL ABILITY
READING ABILITY
REASONING ABILITY
SOCIAL ABILITY
SPATIAL ABILITY
VERBAL ABILITY

CREATIVITY
IMAGINATION
INGENUITY
INVENTIVENESS
ORIGINALITY
RESOURCEFULNESS

ELIGIBILITY
EMPLOYABILITY
QUALIFICATIONS
```

Permuted List of Descriptors
IRT Part 3

```
           ABILITY
           ABILITY GROUPING
           ABILITY TEST
   GENERAL ABILITY
 LISTENING ABILITY
MECHANICAL ABILITY
    MENTAL ABILITY
   MUSICAL ABILITY
 NUMERICAL ABILITY
   READING ABILITY
 REASONING ABILITY
    SOCIAL ABILITY
   SPATIAL ABILITY
    VERBAL ABILITY
```

@ after a related term (RT) = reciprocal related term

For the meaning of SN * see text

The results were extremely satisfying. The German PI vocabulary comprises approximately 800 words, of which 421 are key index terms. Of these 24 are countries for which no place is allotted in the faceted array (thus, perhaps the necessity of creating an 18th facet for geographical areas). The remaining 397 key words fell into four groups:

- German terms which had a direct equivalent when translated into English, e.g. Testentwicklung = test construction; Leistungstest = achievement test, ca. 49%.
- German terms which were specialised forms of the IRT general terms with, however, the same word root when translated, e.g. Automatische Sprachübersetzung -translation; Bildungsplanung - planning, ca. 24%.
- German terms for which there was no direct equivalent in the English vocabulary yet for which there were corresponding terms peculiar to the American educational system, including school types (e.g. *Realschule*) degrees or tests, etc., ca. 24%. This group included those terms which were used in a grammatical form other than that utilised for their equivalents in English, necessitating placement in another (generally obvious) facet (e.g. *'Lernhilfe'* denotes the prompt or cue itself, whereas the IRT employs the terms 'prompting' or 'cueing' as activities); as well as those new terms without an English equivalent or root stem, yet which nevertheless permitted a fairly well-defined semantic placement (e.g. *Prozentrechnen, Algorithmus, Unterrichtsmodell*).
- Those German terms for which there was no provision in the facets and which required modification or extension of the facets for their placement, e.g. *Forschung, Entwicklungsland, Industrie,* ca. 3%.

Although these figures are approximations and do not include reference terms, they do permit a rough estimation of the problems one can expect to encounter when attempting to incorporate the whole of the educational terminology into this form. Terms in the first two groups, ca. 73%, could fairly easily be fitted in by determining general semantic and grammatical equivalence. Such determinations can become extremely problematical in certain cases, however, owing to the difficulties caused by subtleties in language usage, which moreover cannot be controlled in both languages. With few exceptions, most terms in the third group could be clearly placed by expanding facet groups. Only approximately 3% require a possible structural change in the faceting. Were one thus dealing with several languages to be integrated into the faceted array form, approximately 25% of the terms might be subjected to variation in placement owing to either different grammatical and semantic usage or total lack of provision for the terms or their near equivalents in the faceted array form.

These general results were positive. The faceted array form proved to be sufficiently comprehensive and flexible to incorporate almost all PI terms without distortion of meaning.

5.32 *Problems of polyglot thesauri illustrated by means of an example: Facet 15 of the IRT (English-German)*

Appendix 2 shows the General Headings of one German facet (15: Sachen, Tiere) in comparison with the same facet of the IRT (28).

---

(28) *EDITOR'S NOTE:* Space does not permit us to reproduce the complete list of equivalent German and English terms given under the 29 sub-facets of facet 15, as provided by the author. Interested readers should apply to the Pädagogisches Zentrum, Berlin.

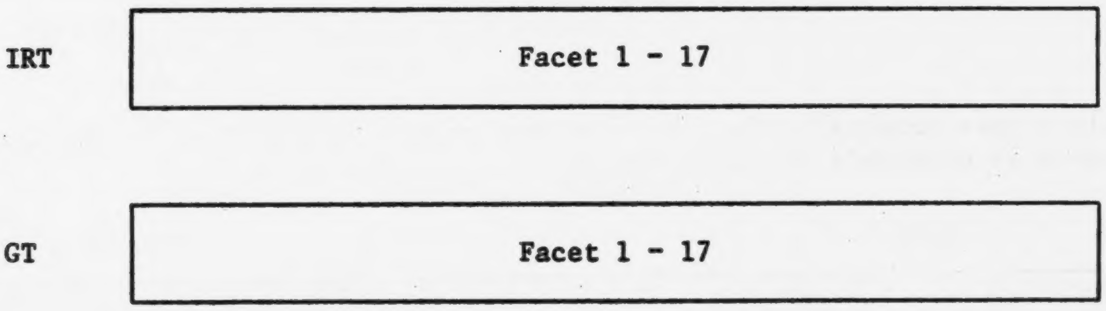
It should be noticed that:

- the German terminology is more specific than the American vocabulary,
- the German terminology is not complete,
- a translation of the IRT was not the aim.

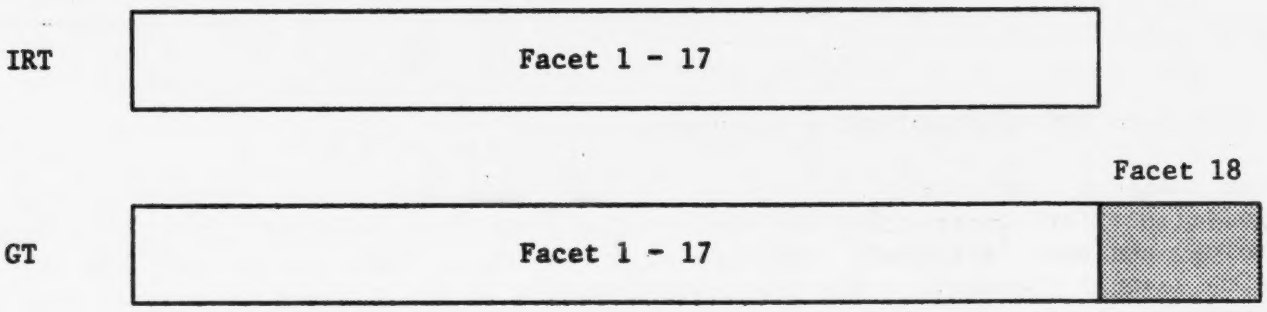
5.33 *Semantic and systematic difficulties*

The following diagrams illustrate four main difficulties which arise from semantic and/or systematic problems. The comparison is based on the IRT and a projected German thesaurus (GT) for the field of education.

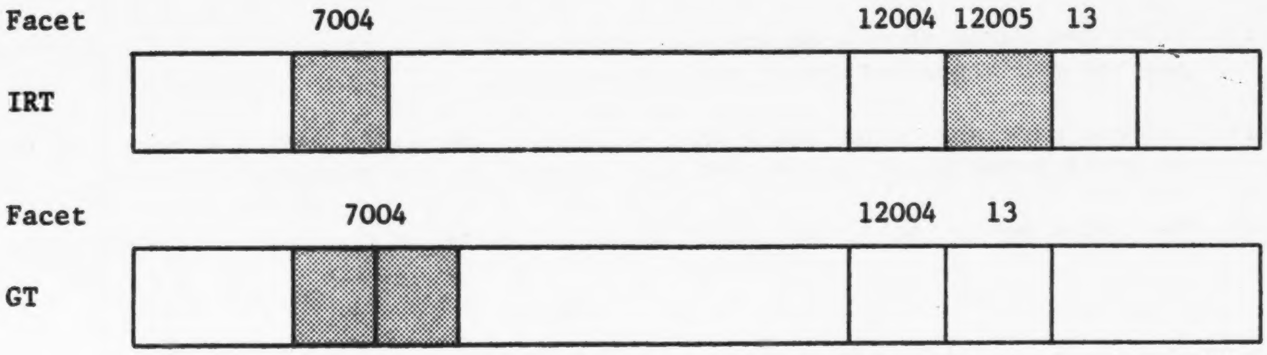
The IRT and the German thesaurus are congruent in theory, thus:


IRT
Facet 1 - 17
GT
Facet 1 - 17

but a simple systematic difficulty gives rise to the necessity of creating an eighteenth facet for proper names:

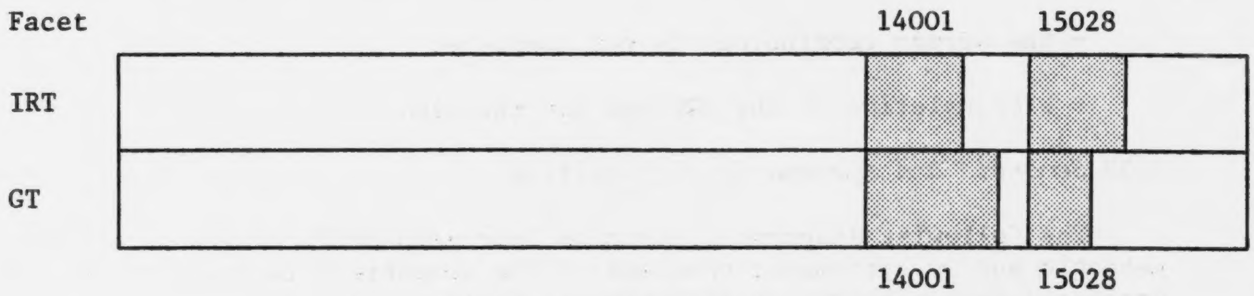

IRT
Facet 1 - 17
Facet 18
GT
Facet 1 - 17

and when systematic and semantic difficulties combine, either a modification to the German thesaurus is desirable:


Facet
7004
12004
12005
13
IRT
Facet
7004
12004
13
GT

IRT: 7004: Ideology; 12005: Religious and political affiliation

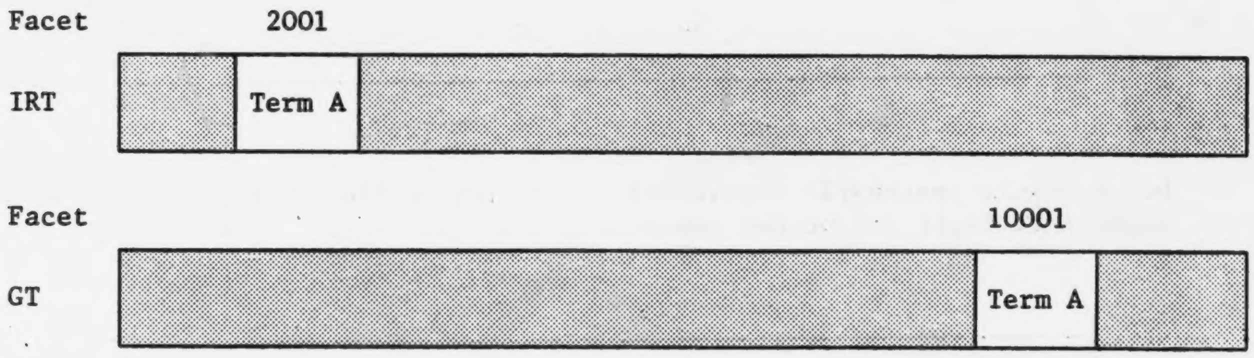
GT : 7004: Ideologie (religiöse und politische Bindung) or a modification to the German thesaurus is necessary:



IRT: 14001: Grade placement level (Facet 14: Educational level)
15028: Schools by special types

GT : 14001: Allgemeine Schulebenen
15028: Schultypen

finally simple semantic difficulty can mean that a modification to the German thesaurus is necessary (no alternative):



6. GUIDELINES FOR CONSTRUCTING A THESAURUS (29)

The American *Committee on Scientific and Technical Information* (COSATI) published in 1967 guidelines for constructing thesauri. The guidelines refer mainly to questions of terminology and more 'external' thesaurus organisation. They are derived from practical work with several thesauri. The proposals embodied in the guidelines are, at many points, of a general nature and consequently lack binding force (relatively speaking). The purpose of the guidelines is to induce homogeneity, at least with respect to the 'external' form of thesauri, and even if diverse special fields are affected.

A thesaurus is expected to fulfil the following requirements:

(i) Every thesaurus should be preceded by an introductory note describing the exact purpose and organisation of the thesaurus.

(ii) Extreme care should be taken when selecting the index words, especially so as to avoid terms which are either too general or too specific.

(iii) The index words should be either nouns or words in the form of nouns.

---

(29) "Guidelines for the Development . . ." (see bibliography).

(iv) In general the plural form should be used. (This requirement cannot apply to the German language or terminology in the educational field - there are many singular nouns which have no plural form.)

(v) Any ambiguity of index words must be eliminated by means of additional explanations:

- (a) by forming compound nouns;
- (b) by using clarifying additional words; for example, in the case of homographs: MERCURY (messenger of the gods), MERCURY (planet), MERCURY (metallic element);
- (c) by using explanations (scope notes); but extensive lexical definitions should not be the aim.

(vi) A direct word order. The adjective should be inverted only in special cases (to make the homogeneousness within a word-group clear).

(vii) In the case of synonyms *one* term acts as a 'subject index term' to which other synonyms are referred ('see'). Moreover, it is useful to treat related words in the same fashion as synonyms, as well as words which are related by virtue of the fact that their meanings are opposite (e.g. 'stupidity', see also 'cleverness').

(viii) Punctuation marks should, if possible, be avoided. "Punctuation marks should be used in subject index terms only when needed since they complicate filing and are difficult to use consistently."

(ix) Abbreviations should be used only if intelligibility does not suffer as a result.

(x) The following reference words have been found from previous experience to be particularly practicable:

	Abbreviation
use	*use*
used for	*UF*
broader term	*BT*
narrower term	*NT*
related term	*RT*

(Perhaps it would be worthwhile using the English notations for a German thesaurus too.)

(xi) The reference words 'broader term' and 'narrower term' should always be used reversibly.

7. PRE-REQUISITES FOR THE CONSTRUCTION OF A THESAURUS

7.1 *The potential users of the thesaurus*

Up to now the potential users of a future thesaurus have been implicitly assumed in many cases. It would be necessary to question them as to their requirements and, in this way, test the extent to which the conceptions and possibilities of the authors and user needs can be matched.

One group of potential users would comprise education specialists and students at universities and colleges of education, another would be the teachers, and then of course there are the workers in the field of educational documentation; the latter should find a conceptual situation which enables documents to be indexed correctly and research work to be carried out as effectively as possible.

7.2 *Specific thesauri or a comprehensive thesaurus?*

The issue of specifying terminology hangs together with the following question: a comprehensive thesaurus for 'education' or individual specific thesauri? The demand for a specification of the terminology is dominant. This is the starting-point, and it is secondary whether the separate specific thesauri are published as a single volume or as individual volumes, co-ordinated at the same time by a 'supra-thesaurus' (*Dachthesaurus*). This is a question of scope and practicability. American models (especially the IRT) appear to indicate that *one* thesaurus is by all means possible and even desirable.

7.3 *Stages in thesaurus construction*

(Starting from the assumption that working out a taxonomy is an indispensable requirement for the construction of a thesaurus.)

(i) Checking terminological material which may be available, or, if no such material is available, collecting it.

(ii) Classifying the material alphabetically and the first compilation of related concepts.

(iii) Producing concept lists or word domains for smaller areas.

(iv) Producing a comprehensive taxonomy.

(v) Completion of the taxonomy and the alphabetical index.

(vi) Testing the thesaurus thus constructed by using practical experimentation.

(vii) Printing the thesaurus.

7.4 *Time and cost involved*

Concrete details regarding time and costs involved could not be obtained as specifically as desired. All the details, with one exception, refer to thesauri taken from the natural sciences or technological fields. Broadly it can be said that as a rule a thesaurus does not need more than three years for its construction and that a permanent editing team of three can be regarded as a minimum for a fair-sized specialist field. It has become generally accepted that the thesaurus should be completed as rapidly as possible "to guarantee a degree of continuity in the composition of the work-group as well as a continually close contact with the vocabulary to be processed" (30).

The production of the *Thesaurus of Engineering Terms* was, on the contrary, a mammoth operation of the following dimensions: 131 workers selected 10,515 terms over a period of 27 weeks' full-time work.

Time required for the completion of other thesauri:

- *Thesaurus of ERIC Descriptors*, USA, (Education): 9 months (only a section of the educational field).

---

(30) SOERGEL D: "Klassifikationssysteme und Thesauri" (see bibliography).

- Thesaurus *Holztechnik* (DDR): 2 years (*The Technology of Wood,* GDR)
- *A Polytechnic Thesaurus* (Norway): 3 years.

7.5 *Constructing one's own thesaurus (with the aid of the IRT)*

The author of the present study proposes the use of the IRT (constructed over a working period of many years) as a standard. The work stages would then be as follows:

(i) If terminological material is available: classifying the terminology with the aid of the IRT faceted array; if no terminological material is available: translating specific areas from the IRT into the respective language (in consultation with specialists).

(ii) Correction of the taxonomy (French, German etc.) or of the translation by appropriate specialists.

(iii) Co-ordinating the separate areas in the thesaurus (alphabetical index, faceted array).

(iv) Testing the thesaurus by practical experimentation.

(v) Printing the thesaurus.

APPENDIX 1

Thesaurus of ERIC Descriptors - Translation from English into German

Examples from the list proposed by the *Dokumentationsring Pädagogik* (DRP)

Abbreviations:

D: = Descriptor
NT: = Narrower Term
BT: = Broader Term — ERIC Thesaurus
RT: = Related Term
UF: = used for
use

T: = Translation
DRP: = List of terms of the DRP
s. = see — list of terms of the DRP
sa. = see also
--- = no term in the list of the DRP

D:	ability
T:	Fähigkeit, Begabung
DRP:	Fähigkeiten sa. Begabung Begabung

UF:	Low ability
T:	Minder-, Schwachbegabung
DRP:	Begabung sa. Dummheit

NT:	academic ability
T:	akademische Fähigkeit
DRP:	---

NT:	cognitive ability
T:	Erkenntnisfähigkeit
DRP:	---

NT:	language ability
T:	Sprachfähigkeit; Fähigkeit zum Sprachenlernen
DRP:	---

NT:	predictive ability (testing)
T:	Aussagefähigkeit
DRP:	---

NT:	verbal ability
T:	verbale Fähigkeit
DRP:	---

RT:	ability grouping
T:	Begabungsgruppierung
DRP:	---

RT: ability identification
T: Fähigkeitsbestimmung (Fähigkeitsfeststellung)
DRP: Begabtenauslese
s. 1. Begabung:Auslese
2. Auslese:Begabung
sa. Auslese
Elite
Schülerauslese
Studentenauslese

RT: achievement
T: Leistung
DRP: Begabung
sa. Leistung

RT: aptitude
T: Befähigung
DRP: Begabung
sa. Anlagen
Eignung

RT: aspiration
T: (Leistungsverlangen)
DRP: ---

RT: talented students
T: begabte Studenten
DRP: Begabtenauslese
sa. Studentenauslese

RT: handicapped
T: benachteiligt
DRP: --- (benachteiligte Kinder)

RT: performance
T: Leistung
DRP: Begabung
sa. Leistung

RT: talent
T: Begabung
DRP: Begabung
sa. Talent

D: ability grouping
T: Fähigkeitsgruppierung
DRP: ---

BT: student grouping
T: Studentengruppierung
DRP: ---

RT: identification test
T: Bestimmungstest
DRP: (Begabungstest; Eignungsprüfung)

APPENDIX 2

Comparison between the General Headings of one facet of the Information Retrieval Thesaurus of Education Terms *and their equivalents in a projected German Thesaurus*

IRT General Headings	Facet	German Equivalents
Things	15	Sachen, Tiere
	Sub-Facet	
General terms	15001	Allgemeine Ausdrücke
Visuals, primarily non-projected	15002	Visueller Bereich (Nicht Film und Projektionen)
Charts Graphs Cards	15003	Diagramme, Graphische Darstellungen, Karten
Films and slides	15004	Film, Diapositive
Cameras	15005	Kameras
Projectors and viewers	15006	Projektoren, Bildbetrachter
Television	15007	Fernsehen
Display boards	15008	Tafeln
Teaching machines	15009	Lehr-, Lernmaschinen
Realia and simulators	15010	Modelle, Simulatoren
Specialised training devices	15011	Spezielle Lerngeräte
Measuring devices	15012	Messschreiber
Timing devices	15013	Zeitmessgeräte
Sound recordings Radio Sound system	15014	Rundfunk, Schallplatte
Computers Calculation aids	15015	Computer, Rechenmaschine
Office machines Control panels	15016	Büromaschinen
Toys	15017	Spielzeug
Conveyances	15018	Transportmittel
Weapons	15019	Waffen
Furniture Parts of building	15020	Gebäudeteile, Möbel
Lenses	15021	Optische Linsen
Materials	15022	Materialien, Verbrauchsgüter
Marking devices	15023	Schreibmaterialien
Musical instruments	15024	Musikinstrumente
Animals	15025	Tiere
Books Printed materials	15026	Bücher, Gedrucktes
Buildings Rooms Facilities	15027	Gebäude, Räume
Schools by special types Social and medical institutions Legal and penal institutions	15028	Schultypen - . Soziale und medizinische Einrichtungen
General locations Administrative divisions	15029	Allgemeine Ortsbezeichnungen

BIBLIOGRAPHY

ANIS. Allgemeines Nicht-Numerisches Informationssystem für IBM 360. Einführung. IBM Deutschland (1968).

BAUER G: Anwendung des Prinzips der Facettenklassifikation für den Aufbau von Thesauri. Erläutert am Beispiel "Thesaurussystem Chemie". In: ZIID-Zeitschrift, 14 (1967) 3, p. 72-85.

BAUER G: Der Grundthesaurus "Chemie" - Probleme des Aufbaus und der Ergänzung durch Fachthesauri. In: Symposium der Mitgliedsländer des RGW. Komplexe Mechanisierung und Automatisierung. Moskau, 10.-14. Juni 1965. Berlin: ZIID-Schriftenreihe 7, p. 3-33.

BAUER G: Die Bedeutung der Kategorien als "Ordnungsmittel" in Thesauri und Speichern der elektronischen Datenverarbeitungsanlagen. In: ZIID-Zeitschrift, 15 (1968) 2, p. 61-65.

BAUER H: Aufstellung eines Thesaurus für Elektrotechnik mit Hilfe von Maschinenlochkarten. In: Nachrichten für Dokumentation, 18 (1967) 1, p. 6-12.

BLANKENSTEIN G: Die hierarchische Struktur eines Thesaurus und die Eindeutigkeit der Deskriptoren. In: ZIID-Zeitschrift, 14 (1967) 4, p. 110-113.

BLANKENSTEIN, G: Erarbeitung von Thesauri unter Berücksichtigung der zukünftigen Entwicklung. In: ZIID-Zeitschrift, 14 (1967) 7, p. 177-180.

DANNER H: Aufbau einer Begriffsliste für die Dokumentation des Standes der Technik auf chemischen Teilgebieten. In: Nachrichten für Dokumentation, 16 (1965) 2, p. 68-75.

DOBROWOLSKI Z: Analyse der Klassifikationssysteme. In: Nachrichten für Dokumentation, 16 (1965) 3, p. 142-149.

DÜRR H G: ANIS - ein allgemeines nichtnumerisches Informationssystem für das IBM-System 360. In: IBM-Nachrichten, 18 (1968) 190, p. 284-290.

FOSKETT D J: Classification and Indexing in the Social Sciences. London: Butterworth (1963) 190 p.

Guidelines for the Development of Information Retrieval Thesauri. Washington, D.C.: Committee on Scientific and Technical Information (1967).

HALLIDAY M A K: The Linguistic Basis of a Mechanical Thesaurus. In: Mechanical Translation, 3 (1956) 2, p. 81-88.

HOFFMAN G: Zur Vereinbarkeit von Fachthesauri für das Patentschrifttum. In: ZIID-Zeitschrift, 15 (1968) 5, p. 215-220.

HUNGERMANN E H: Zur Planung, Erstellung und Bewertung von Dokumentationssystemen. In: DFW (Dokumentation, Fachbibliothek, Werksbücherei), 12 (1964) 5, p. 147-149, 151-158, 160-167.

KÜRSCHNER G: Hinweise zur Bildung von Thesauri für die Informationstätigkeit. In: ZIID-Zeitschrift, 13 (1966) 2, p. 41-42.

LONDON G: A classed thesaurus as an intermediary between textual, indexing and searching languages. In: Review of International Documentation, 32 (1965) 4, p. 145-149.

LUHN H P: Auto-Encoding of Documents for Information Retrieval System. In: Modern Trends in Documentation (Ed. Martha Boaz) London, New York: Pergamon Press (1959), p. 45-58.

MASTERMANN M M: The Thesaurus in Syntax and Semantics. In: Mechanical Translation, 4 (1957) 1-2, p. 35-43.

MEYER-UHLENRIED K H: Klassifikation und automatische Dokumentation. In: Nachrichten für Dokumentation, 16 (1965) 2, p. 61-67.

MEYER-UHLENRIED K H: Ordnungsprinzipien der Speicherung in der Dokumentation. In: Nachrichten für Dokumentation, 16 (1965) 1, p. 18-22.

NEDEL F: Thesaurus in der Dokumentation. Vielschichtigkeit des Begriffes und historische Entwicklung. In: Nachrichten für Dokumentation, 17 (1966) 1-2, p. 5-13.

Nationales Symposium. Thesauri für mechanisierte Informationsrecherchesysteme. Berlin, 3.-6.4.1967. Berlin: Zentralinstitut für Information und Dokumentation (1967) 279 p. ZIID-Schriftenreihe 16.

OPIR K D: A Root-Discriminating Thesaurus. In: Information Storage and Retrieval, (1964) 4, p. 239.

PIETSCH E H: Die zukünftige Entwicklung der Dokumentation. In: Libri, 12 (1963) 4, p. 287-319.

PRESPERSKI H and ZURAWSKI K J: Der Thesaurus als Grundlage für die Sachregisterbearbeitung bei technischen Referatorganen. In: Nachrichten für Dokumentation, 17 (1966) 1-2, p. 19-22.

ROLOFF H: Lehrbuch der Sachkatalogisierung. Leipzig (1950), esp. p. 83-108.

ROTHKIRCH-TRACH K-C and NELTE von: Erfahrungen und Erkenntnisse bei der Aufstellung von Thesauren. In: Nachrichten für Dokumentation, 15 (1964) 3, p. 118-121.

SCHEELE M: Ein Verfahren zur automatischen Klassifizierung für Veröffentlichungswesen, Bibliothekswesen und Dokumentation. In: Die Naturwissenschaften, 52 (1965) 1, p. 1-10.

SCHEELE M: Thesaurus - Baustein jeder Fachdokumentation. In: Nachrichten für Dokumentation, 15 (1964) 1, p. 1-4.

SILBERNAGEL C: Anwendung der Logik auf die Lösung von Thesaurusproblemen. I : ZIID-Zeitschrift, 14 (1967) 3, p. 140-142.

SOERGEL D: Klassifikationssysteme und Thesauri. Frankfurt a. Main: Deutsche Gesellschaft für Dokumentation 1969, 173 p.

TAUBE M: Extensive Relations as the Necessary Condition for the Significance of "Thesauri" for Mechanized Indexing. In: Journal of Chemical Documentation, 3 (1963), p. 177-180.

Technische Thesauri in Westeuropa. (von Hans-Dieter Rosenbaum). In: ZIID-Zeitschrift, 13 (1966) 4, p. 114-116.

VERSUČENKO Ju G P: Anwendung eines permutierten Registers zur Analyse und Vervollkommnung von Thesauri. In: ZIID-Zeitschrift, 15 (1968) 5, p. 206-209. (Übersetzung aus dem Russischen).

VICKERY B C: Classification and Indexing in Science. London: Butterworth (1959) 235 p. (2.Aufl.).

VICKERY B C: Faceted Classification Schemes. New Jersey (1966). Rutgers Series in Systems for the Intellectual Organisation of Information Vol. V.

VICKERY B C: A Retrieval System Theory. London: Butterworth (1965) 191 p.

WAGNER S: Automatische Stichwortanalyse nach dem Rangkriterienverfahren. Diss. Karlsruhe 1966.

WALL E: Information Retrieval Thesauri. New York: Engineers Joint Council (1962).

Zur Thesaurusfrage. Arbeitsunterlagen. Frankfurt a. Main: Deutsche Gesellschaft für Dokumentation (1966).

Unpublished Literature

BAUER G: Alphabetische Liste der wichtigsten Termini zum Begriff "Deskriptor". Berlin: Deutsche Akademie der Wissenschaften, Wiss. Redaktion der Zentralblätter, Abt. Forschung und Entwicklung (1966).

BRENNER E H: Thesaurus construction - historical background and use considerations. Vortrag: FID-Kongress, Washington, Oktober 1965.

CAMPBELL D J: Analytical techniques for the radical revision of thesauri. Vortrag: FID-Kongress, Washington, Oktober 1965.

WÄHLIN E: A general analysis of different types of classification systems characterised by their objects. Vortrag: FID-Kongress, Washington, Oktober 1965.

COMPUTER HANDLING OF SOCIAL SCIENCE TERMS AND THEIR RELATIONSHIPS

by

E J COATES

P.A/53

The BTI indexing system 55

Possible application of the BTI system to the field of educational sociology 63

Specimen index 70

ACKNOWLEDGEMENT

The writer's thanks are due to Peter F. Broxis FLA for useful discussions on relational subject analysis applied to the literature of educational sociology.

I THE BTI INDEXING SYSTEM

This study begins with a consideration of some general parameters of indexing systems. These are enumerated in the following series of questions.

(1) Is intellectual work applied mainly at the input of the system in order to reduce the need for intellectual work by the searcher, or vice versa?

(2) If intellectual work is applied at the input, is this limited to the first phase, comprising summarisation of the overall content of a given documentary unit?

(3) Is the second phase of intellectual work, comprising vocabulary control, practised?

(4) Is the third phase of intellectual work, comprising syntactic organisation of the indexing language, practised?

(5) Is the fourth phase of intellectual work, comprising hierarchical organisation of the index language, practised?

(6) Is the organisation designed for searchers associated with one or many disciplines?

(7) Is the organisation designed for narrow sector or broad field search?

These parameters are now discussed with special reference to the *British Technology Index* system.

Since the advent of computers much progress has been made in the production of indexes without the application of intellectual effort. Such indexes are the product of the manipulation of words which may be taken from the titles or texts of documents, either directly or after being subjected to an automatic selection process derived from statistical considerations. In either case the assumption is made that without intellectual effort at the output, only comparatively low standards of relevance and retrieval precision are obtainable.

Although *British Technology Index* is a mechanised system, it does not form part of the trend mentioned above. In the *British Technology Index* system intellectual effort is applied at the input with the object of minimising the need for such effort by the searcher at the output. This application of intellectual effort at the input is an important factor in the costs of producing an index; it amounts to perhaps 25% of all costs of a printed index. It may however be possible to justify the cost in terms of social accounting. It is likely to be more economical to carry out the intellectual work once at the input than to perform it many times - and repeatedly, each time the output is consulted by a searcher. On the wider view of the social costs of failure of information transfer, the case for applying intellectual effort at the input appears to be formidable. The indexer is almost by definition a person trained to make the necessary intellectual effort. The searcher is probably not so trained. His understanding of the effort required is unlikely to be systematic and may often be absent altogether, and the index may fail as a system precisely for this reason. This argument is not, of course, applicable where there is a trained intermediary between the index and the searcher. It is not applicable to an index intended to be consulted only by librarians or information specialists. It is however, consonant with the intention that *British Technology Index* should be usable directly by the ultimate consumer, the applied scientist or engineer who needs information.

The primary intellectual work applied to *British Technology Index* consists of the summarisation of the content of a technical paper or article. It is important to distinguish summarisation of the content from enumeration of the content. Summarisation involves the subordination of enumerated concepts under more general inclusive heads. Thus in the context of metalworking, the enumerated concepts, "selection of material", "shaping", "finishing" and "testing", are summarised under the inclusive concept "manufacture".

Summarisation is thus an economising device operating within the limits of a chosen document unit, which is the whole paper in the case of *British Technology Index*, but which could, with equal consistency, be the main section of a paper, the paragraph, or the sentence. It is necessary to draw attention to the proviso above: within the limits of a chosen document unit. Although summarisation is the result of "generalising" the various contributory concepts found in the paper, it does not imply that the overall or comprehensive concept for the whole paper, thus established, will be similarly "generalised". A complete paper on "penetrant flaw testing" - a concept which perhaps figured in the "testing" section of the earlier hypothetical paper on metal manufacture - is entered under this heading and not under "Testing". Summarisation implies no departure from the concrete or the specific at the level of the chosen document unit. One is thus brought to a consideration of the relation between specificity in indexing and exhaustivity. The latter depends upon the particular document unit within which it is decided to summarise. Specificity depends upon the rigour with which generalisation is inhibited at the level of the chosen document unit. *British Technology Index*'s exhaustivity is a function of the choice of the whole paper as the unit to be summarised. Its specificity is the measure of precision with which the index heading corresponds to the summarised idea-content of the whole paper. Summarisation inevitably involves consideration of concepts. It necessarily constitutes a step towards concept indexing and away from indexing as word manipulation.

The secondary intellectual work in concept indexing is the establishment of an index language, distinguished from natural language by the fact that each word or index language unit is correlated with a concept in such a way that for every concept there is one, and one only index language unit. Conversely for each index language unit there is one and one only corresponding concept. Thus the synonyms, alternatively ordered compound expressions, and the homographs of natural language are eliminated. The process of setting up and maintaining such an index language and applying it to the description of documents has been termed vocabulary control. In practice, vocabulary control as here defined, has often been associated with severely restricted vocabularies. It has been extended beyond the elimination of conceptually equivalent terms in the index language to the merging of conceptually related terms. This merging is not compatible with specificity. Specific entry is required for *British Technology Index*: therefore, in the *British Technology Index* system, vocabulary control is concerned only with the elimination of possible conceptual equivalents in the indexing language. If a concept is by definition distinct, even only slightly, from another concept, it will have its separate indexing language correlate. The *British Technology Index* vocabulary is thus large as controlled vocabularies go, increasing approximately by 1500 per month if all new term combinations are counted, or by 50 per month if the appearances of new terms alone are counted. Some complex economic considerations come into play in connection with this controlled vocabulary In the indexing process itself, the large vocabulary can be memorised only partially Every tentative indexing decision has to be looked up in an authority file of precedents. This authority file itself has been a considerable load on costs owing to the effort required to keep it updated in a continuously developing situation. On the other hand, as any separately definable concept may have its index term, there is probably less decision time involved in assigning an index term in a vocabulary of this kind, than in a more drastically reduced, and temporarily closed ended vocabulary. The burden of maintaining a manual authority file was a strong motivating factor in the *British Technology Index* mechanisation program Since April 1968, a computer produced authority file, a byproduct from other data processing operations, has been in use for *British Technology Index*. The main advantages of the computer produced authority file are that there is no continuous maintenance required; production of multiple copies in computer print out form is easy, and each indexer can have a relatively compact copy on his own desk. Updating is automatically carried out each month The disadvantage of the computer print out authority file is that the optimum periodfor which it can be cumulated appears to be six months. This is determined partly by thesharply increased cost of computer storage and sorting effort beyond six months, and partly by the fact that errors in the authority file are expensive to correct systematically and at the same time dangerous to leave. The error rate is such that we feel that the retention of errors in the authority file is tolerable for six months but no longer. The six monthly cut off and re-start of the authority file means that for part of the year the indexers have to consult two authority files - at the end of the year the earlier of the two files will be superseded by the printed annual cumulation of the *Index*. The manual authority file corresponding to the first six years' intake of material is retained for the time being as static backing store.

The central enigma in attempting to store information in an index is that on the one hand the user will question the index by single words or phrases, whereas a subject heading which attempts to give the subject of a technical paper with reasonable precision will usually require several words to indicate the intersection of concepts denoted by single words. Most articles treat, or purport to treat of new knowledge, which proceeds by the intersection of two or more "established" ideas. This dilemma has produced a number of dead-end consequences in the development of indexing methodology. It has produced a convention that subject index headings must consist of single words or phrases, and the corollary that works on contemporary developments in knowledge, which cannot be specified in a single word or phrase, cannot be indexed specifically.

The notion of logical intersection as it occurs in the conjunction of simple terms for specifying terminologically composite concepts, needs to be examined concretely. There are several kinds of intersection. The symbolism corresponding to a particular kind of intersection may convey indispensible information in a subject heading. The various kinds of intersections are easily illustrated by examples from the field of technology:-

(1) FIBRES: Extrusion

(2) FIBRES: Anisotropy

(3) FIBRES: Inorganic

The relations between Fibres and the second concept are different in each case. In example (1) the relationship is that between an action and what is acted upon. In example (2) the relation is that of a property belonging to an object. In example (3) we observe a relationship in which the second term delimits a species of the first term. These differences become of great importance where the computer programmer is asked to devise an algorithm for manipulating the order of elements in composite subject headings. As the foregoing attempts to define particular relationships suggest, it is impossible to take relationships into consideration without, at the same time, categorising concepts. The definitions speak of actions, properties, objects, and so on. An approach to relationships via categorisation is illustrated from a series of terms in a sequence randomly selected from a *British Technology Index* annual volume:-

IONOSPHERE

CURRENT

RESEARCH

ROCKETS

ELECTRONS

COLLISIONS

RADAR

DENSITY

MEASUREMENT

SATELLITES

ARTIFICIAL

X-RAYS

COUNTERS

PROPORTIONAL

TOPSIDE

SOUNDING

INSTRUMENTS

PHOTOGRAPHIC

EMULSIONS

SENSITIVITY

CALIBRATION

If we categorise as many as possible of these terms into THINGS and ACTIONS we have:-

THINGS	ACTIONS	DOUBTFUL	OTHER
IONOSPHERE	RESEARCH	SOUNDING	ARTIFICIAL
ROCKETS	CALIBRATION	CURRENT	PROPORTIONAL
RADAR	COLLISIONS		PHOTOGRAPHIC
SATELLITES	MEASUREMENTS		SENSITIVITY
X-RAYS			DENSITY
COUNTERS			
TOPSIDE			
INSTRUMENTS			
EMULSIONS			

The items under "doubtful" illustrate the impossibility of making this categorisation by formal tagging of parts of speech characteristics. 'Sounding' may be an adjective meaning 'creating' sound or 'probing a depth', a gerund meaning 'for creating sound' or 'for probing a depth', or a noun denoting the action of sounding in either of its two meanings. Let us assume, however, special knowledge of the contextual meaning and assume that 'current' means a flow of electrons, and sounding means the action of probing at depth or height. We thus assign both terms to the action category. Turning now to the "other" column we observe that Sensitivity and Density are members of a further category, Properties. As far as Artificial, Proportional, Photographic are concerned, these do not fall into any category, yet their relational function is in clear distinction from that of all other terms categorised. Whereas the categories are defined semantically, each consisting of terms linked in meaning, the members of this final group have no semantic common ground, even of the most generalised sort. They may be categorised *only* in terms of their relational role with respect to terms in the earlier categories, namely, Things, Actions, Properties. This relation is that they indicate the types or species of the various Things, Actions and Properties. Thus the term "Artificial" used in conjunction with "Satellites" designates a sub-set of all satellites.

This type-specifying category is necessarily defined in relational terms. It might seem at first sight that the semantic criteria which define the remaining categories labelled Things, Actions, Properties are of a fundamentally different kind. It appears however, when we examine the matter more closely that the semantic criteria are useful provisional devices which, however, advance analysis only a short distance. The Thing-Action combination INSTRUMENTS RESEARCH could mean either Research upon instruments or Instruments as tools for Research. In other words, we need to add the idea of direction and hence of relation between Thing and Action.

Instruments ⟵ Research is clearly distinguishable from

Instruments ⟶ Research

Similar directional problems occur in the combination of Actions and Properties

Measurement Sensitivity

can mean

Measurement ⟵ Sensitivity Sensitivity of measurement

or Measurement ⟶ Sensitivity Measurement of sensitivity

Finally terms from the same category are capable of combination

e.g. Ionosphere	Topside	Topside of the ionosphere.

We can designate this as a Part, but this is a purely relational designation.

Syntactic organisation of index terms appears therefore to be primarily a question of identifying relationships between terms. These relationships will correspond with such semantic categories as Thing, Action, Property and may be tentatively named by reference to the categories, but the categories, themselves, are not an entirely firm base for an indexing syntax. Many terms can appear in different semantic categories in different relational contexts. In particular, many substantive terms are capable of an adjectival role in a type-specifying relation, as for instance

Satellites/Research	Research on Satellites
Satellites/Research	Research Satellites

In natural language the distinctions which have emerged in the foregoing discussion are signified by the use of prepositions or inflectional endings or both. English relies almost exclusively on prepositions without inflexions and the *position* of the elements in a sentence, is pressed into service for the indication of relationships. This is a phenomenon of great importance for indexing languages. Up to 1968 the *British Technology Index* relied entirely upon positional indication of term relationship, but subsequently for the purpose of specifying desired computer manipulations, a relation notation using punctuation symbols was brought into use.

The full list of syntactic relations used in *British Technology Index* is as follows. The table indicates the positional rules and the relation indicating punctuation used in headings.

		Example	*Meaning in natural language*
1	TYPE SPECIFYING RELATIONS Thing, Type (by reference to a Thing, Action or Property)	Satellites, Research	Research satellites
	Thing; Type (by reference to material)	Beams; Aluminium	Beams *made of* aluminium
	Material, Type (by reference to action or property)	Aluminium, Anodised	Anodised aluminium
	Action, Type (by reference to action or property)	Moulding, Injection Collisions, Inelastic	Injection moulding Inelastic collisions
	Property, Type (by reference to property)	Velocity, Instantaneous	Instantaneous velocity
	Purpose: Thing (by reference to action)	Printing: Inks	Inks *for* printing
	Thing: ancillary thing	Buses: Garages	Garages *for* buses
2	WHOLE PART RELATION Thing: Thing (whole) (part)	Buses: Engines	Engines *of* buses

	Example	*Meaning in natural language*
3 INTERACTIVE AND ATTRIBUTIVE RELATIONS		
Thing: Action on	Satellites:Research	Research applied to satellites
Thing: Action by	Lasers: Welding	Welding by lasers
Thing: Material	Structures: Steel	Steel *of* structures
Thing: Property	Beams: Thickness	Thickness *of* beams
Action: Action	Packaging: Labelling	Labelling *for* packaging
Action : Property	Collisions : Inelasticity	Inelasticity *of* collisions
Property : Property	Hardness : Ratio	Ratio *of* hardness

It is to be noticed that in English the prepositions used in the equivalent statements are a fairly reliable guide to the relationships needing to be distinguished, though they need to be used with caution.

The tabulation above gives the positional rules for pairs of elements only. The rules for larger aggregation are as follows:-

(1) The type specifying relational pairs with , (comma) are regarded as being more tightly bound together than relational pairs with : . The latter cannot be interposed between a Thing and its type term.

(2) A combination of the two first interactive relations given above, i.e. Action on a Thing and the same action by a Thing is rendered in the form

Thing acted upon : Action: Thing Acting

e.g. FOIL : Welding: Lasers

(3) Artefacts without clearly established conventional names are entered by citation of their properties in the order

PURPOSE : Shape : Other properties

e.g. PRESSURE VESSELS, Cylindrical; Steel

(4) When a Thing requires to be specified by several properties, citation order proceeds from the more to the less fundamental properties.

Some representative examples are as follows

PIPES; Steel, Mild: Bulge tests, Hydrostatic

RADAR: Tracking: Aerials: Control systems: Computers: Analysis

SHEETS; Titanium: Plastic deformation: Anisotropy

SENSORY-MOTOR PERCEPTION : Time: Estimation: Proprioception

SOLIDS, Granular: Heat transfer: Heat exchangers, Countercurrent

It should be noticed that these headings are usually equivalent to reversed English natural language statements of the subject, with prepositions substituted for the : (colon). For example the final heading quoted is rendered as an English statement as countercurrent heat exchangers *for* heat transfer *to* or *from* granular solids. *British Technology Index* indexers normally produce a preliminary subject statement of article content in this form before converting it to a subject heading.

The heading derived from intellectual work according to the syntactic plan outlined is the starting point for the clerical task of creating cross-references from significant elements in a heading (other, of course, than the first, which itself determines the position of the entry). Users may need to access the material from one or other of these subordinate terms. The simple, foolproof, but usually uneconomic, method of making this provision is to permute the items fully. The *British Technology Index* uses a device for reversing (or "inverting") the headings to produce cross-references which owes its inspiration to Dr. S. R. Ranganathan's chain procedure for the subject index to a classified index. The basic pattern of inversion cross-references produced from a heading is illustrated by the heading

TANTALUM : Passivation : Nitric acid

which produces cross-references

(1) NITRIC ACID : Passivation : Tantalum

(2) PASSIVATION : Tantalum

It will be noticed that the user looking under PASSIVATION for passivation by nitric acid is not immediately informed of the existence of material on the subject. The "chain procedure" term manipulation provides a mixture of specific and non-specific cross-references. If our user had first looked under NITRIC ACID the information that material on the sought subject was listed would be given immediately.

The basic generalised pattern of cross-references for the subject heading

A-: B-: C- : D- is therefore

D-: C-: B-: A- See A-: B-: C-: D-
C-: B-: A- See A-: B-: C-
B-: A- See A-: B-

Because it is important that *British Technology Index* should give an unequivocal cross-reference for specific types of Things, the foregoing pattern is modified by a superimposed permutation wherever the Thing-Type relationship, symbolised by a comma, occurs. This produces the following pattern:-

Heading A-: B-: C-, D-

Cross-references D-C-: B-: A- See A-: B-: C-, D-
C-, D-: B-: A- See A-: B-: C-, D-
B-: A- See A-: B-

The construction of these cross-references has been carried out by computer for *British Technology Index* since April 1968. Coded headings are punched for this purpose by *British Technology Index* staff and the computer produces a sorted print-out which gives all the required cross-references. The turning round, or inversion, of a composite heading to form an intelligible cross-reference, now given in the forward order of an English noun and preposition statement, is a fairly simple matter provided that certain of the relational bonds between elements can be emphasised at the expense of others. This differentiated emphasis in the cross-reference is achieved by the differentiated punctuation used in the input heading. In the latter of the two generalised patterns of cross-references given above, it will be noticed that the comma between C and D on the input disappears between D and C on the first cross reference. This is an example of the measures taken in the computer program to exhibit in the cross-reference the relative strengths of bond between terms.

A fully coded heading input for computer generation of cross-references, has two punctuation symbols between each element. The first, usually retained on the output, signifies the relation between two adjacent elements. The second indicates to the computer whether or not the following element is significant in the sense of being required as a leading term in a cross-reference, and it may have special transposition functions when placed before a term of very high generality.

An example of a fully coded heading is the following:-

SOLIDS,.Granular:.Heat: /transfer:.Heat: /exchangers,.Countercurrent.

From this data the computer produces

COUNTERCURRENT HEAT EXCHANGERS:Heat transfer:Granular solids. See

SOLIDS,Granular:Heat transfer:Heat exchangers,Countercurrent

HEAT:Exchangers,Countercurrent:Heat transfer:Granular solids. See

SOLIDS,Granular:Heat transfer:Heat exchangers,Countercurrent

HEAT; Transfer : Granular solids. See SOLIDS, Granular :Heat transfer

GRANULAR SOLIDS. See SOLIDS, Granular

The program for inversion reference gives the indexer considerable flexibility - probably too much. Where needed, two elements of the input can be manipulated as one, or conversely a phrase in the input can be transformed to single words separated by punctuation. Where for the purpose of a cross-reference entry word it is desired to change the form of the word given in the input - this may be achieved by appropriate coding. A particular feature of the program is that no attempt was made to carry its development to the point at which it could deal with every possible semantic situation presented to it. About 1 to 1.5% of the inversion cross-references in *British Technology Index* need to be derived from heading data too complex to be handled by the program. These are supplied manually.

The fifth parameter mentioned at the beginning of this study refers to the hierarchical organisation of the index language. *British Technology Index* headings are designed as far as possible to aid specific subject approach and contain no hierarchical expression in their internal structure. However, hierarchical linkage between words used as headings is provided by means of cross-references whereby at broad concept terms the names of related terms comprising species of the broad concept are listed. Thus a user who accesses at a broader term than that representing the subject actually required is assisted. The organisation of the hierarchies so constructed is based partially on the U.D.C. in the broader reaches and partly on other special classification systems in the regions of more detail. Usually no term is allowed to occupy more than 1 hierarchical series, but there are exceptions to this, particularly with respect to single application organic chemicals. These are accessible from generic terms signifying both their type of chemical structure and also their field of application. Accessing via related heading references will of course demand three searching steps, when the search begins with an access term, moves to a 'related heading' term, which is itself the leading term of a cross-reference. Thus the example given previously of the Passivation of tantalum by nitric acid could be accessed via Acids or Corrosion or Metals, as follows:-

ACIDS ⟶ Related Headings NITRIC ACID Inversion Reference ⟶ TANTALUM

CORROSION ⟶ Related Headings PASSIVATION Inversion reference ⟶ TANTALUM

METALS ⟶ Related Headings TANTALUM

The practical production of these related heading references is carried out by computer following the inversion routine. Headings and generated inversion cross-references are first sorted in the computer in one sequence, which is subsequently matched against a store of coupled pairs of terms representing 'related headings' cross-references in reverse.

On matching, the appropriate cross-reference is copied from the store and sorted into the output sequence. It is of course necessary for the content of this matching store to be initially input by the *British Technology Index* staff. This however, is a semi-permanent store and once input, the record will "match" against headings or inversion cross-references at any time.

It remains to add that at the intellectual level the assignment of a new previously unused term to a place under its nearest term is an essential part of vocabulary control. The comparison with other terms already linked to the generic term is the most systematic check that can be devised to ensure that the proposed new term is not a synonym of another already in the file.

The sixth and seventh parameters listed at the beginning of this study require only the briefest discussion. The *British Technology Index* is multi-disciplinary in character, and no single citation or facet order of subject heading elements can be optimal for all of the 50 disciplines concerned. If an index cannot be optimal in this sense, the next best feature that it can possess is predictability. If a single facet order is applied with rigorous consistency, users will fairly easily acquire both understanding and foreknowledge of the indexers' habits and practice. Parameter seven occupies the final position in the list by way of a reminder that the system must be designed with the primary purpose of meeting one particular type of information need. Systems which attempt more than this are rarely satisfactory. *British Technology Index* attempts to answer the need for narrow sector or specific subject enquiries. The particular alphabetical form of index described in this Section of the study derives from this consideration. This alphabetical form would almost certainly not be appropriate if the primary need which we were attempting to satisfy were for broad field search facilities. A classified or systematic structure would better serve the latter need. Nevertheless it is perhaps worth stressing that the underlying relational analysis for subject indication is required irrespective of whether the ultimate realisation is alphabetical or systematic in form.

II POSSIBLE APPLICATION OF THE BTI SYSTEM TO THE FIELD OF EDUCATIONAL SOCIOLOGY

It will be clear that no part of the methodology detailed in Part I is circumscribed by the particular kind of material with which *British Technology Index* deals. Nothing in the system depends upon the fact that the records handled are journal articles and papers in the field of applied science. The same processes of summarisation, vocabulary control, syntactic organisation and hierarchical organisation can be applied to any field.

Nevertheless one particular subject field may demand more or less effort than others, at any given point in the foregoing series of processes. For instance the literature of the Social Sciences abounds in terms on the meaning of which there is not even approximate agreement. Such terms also occur in the literature of technology, but here they are proportionately fewer. The result of this is that the summarisation and vocabulary control phases are likely to be more demanding, in certain respects, in the indexing of the literature of educational sociology than in the indexing of technical literature. Sociological terminology is likely to be relatively more recalcitrant to the transformation required to produce a "concept" language for indexing purposes. It is likely however that the field of sociology contains considerably fewer concepts than the field of applied science. This leads to some mitigation of the indexing task.

Some obvious differences between the two subject fields appear when we consider syntactic organisation. In applied science human purposes are a dominating category of concepts, but human beings as such occur only rarely as concepts which require statement in indexing the literature. Much discussion in technology (as also in pure science) centres on one way interactions between artefacts, between materials, or between artefacts and materials. In such interactions consideration of the spectrum of ordered complexity of the various artefacts and materials in conjunction with the action term itself will often cause the user to take for granted the roles of the artefacts or substances concerned - to guess which is the active factor and which the passive. In the example

GOLD: Mining: Excavators

the action term "mining" itself clearly precludes confusion as to which of the other terms is the subject and which the object. Even with more equivocal action terms everyday understanding of human purposes will often indicate the precise relationship, e.g.

IRON: Production: Fuels

is unlikely to be interpreted as Producing fuels from or by means of iron. The reverse interpretation is overwhelmingly more probable.

Considerations such as these are very much less in evidence in sociological or educational material. The main category of concepts is Kinds of Human being, by age, sex, occupation, economic class, educational status, and so on. The concepts may appear in either an active or passive role. In the example

MIDDLE CLASSES: Domination: Intellectuals

it is impossible to say from this who is dominating whom. More information is required. There are broadly two ways in which the information may be given. The first is to cause meaning to depend upon position in the string of terms. Thus in the example given, we could decide that the passive "party" is that which precedes the action term. The subject heading would then be about the domination of the middle classes by intellectuals. An identical positional mechanism is used in *British Technology Index*. In the example given earlier it is the Iron (passive element) which is produced by means of Fuel (active element). The alternative mechanism, which does not exclude the use of the first at the same time is that of indicating the active or passive role by punctuation diversification. Thus

MIDDLE CLASSES: Domination; Intellectuals

or INTELLECTUALS; Domination: Middle class

can be made to mean the domination of the middle classes by intellectuals, if : carries the passive meaning and ; the active one. This device has not, so far, been found necessary in *British Technology Index*, but it is recommended that it should be used in the documentation of sociology to give support to the positional indication system. In the Specimen Index given on p. 70 the generalised pattern A ; B means that A acts upon B while A : B means that B acts upon A.

The following table gives the operators used in the Specimen Index, together with an indication of their relationship to the *British Technology Index* operators:

A, B .	B indicates a Kind of A	As used in B.T.I.
A; B	A acts upon B	: used in B.T.I.
A: B	A is acted upon by B	As used in B.T.I.
A. B	B is a Property or Part of A	: used in B.T.I.
A,. B	B has the environment A	: used in B.T.I.
A /̲B̄	B is a topic taught, or the Subject of opinion by, at, or to A	: used in B.T.I.
A.. B	B is a place at which A occurs	As used in B.T.I.

A. B A,. B and A/̲B̄ transform to B(A B., A and B̲/̄A when quoted, particularly in cross-references, in the reverse of the logical order used in Simple Subject headings. There is no strict logical necessity to transform these operators, but the transformation is a help to comprehension of the roles of concepts in reversed cross-reference order.

It should be noted that the basic relationships needed to form the framework of a structured record for subject retrieval in educational sociology are the same as those needed to form a similar record for technology. However the relationships are more implicit in the structure of technological knowledge itself than in educational sociology. Consequently, in the latter case the relational notation needs to be rather more explicit than has been found necessary for *British Technology Index*.

Both Education and Sociology are concerned with human beings taking part in action, or as the recipients of action carried on by other human beings. In Education the basic situation is that Educands are recipients of educative action carried on by Educators. From the sociological standpoint, it may be looked upon as a particular instance of Social Group development, the Social group being in this case the community of Educands and Educators. In the facet order chosen for the Specimen Index the social group has been cited first - the action by or upon the group has been cited second. In cases where one group acts upon another, the latter, the passive element is cited first, followed by the activity, followed by the active element. (This position role is exactly analogous to the role in *British Technology Index* that Patient (i.e. the passive element) precedes Action, which in turn precedes Agent).

Communities of Educators and Educands are themselves specified in the Specimen Index by citing first the name of the group of educands, and secondly by the name of the educational level at which the community works. Thus any aspect of secondary education for a particular social group (e.g. Canadians) whether it deals with purely educational matters or with questions primarily of concern to the sociologist or educational sociologist is entered under a heading which begins

CANADIANS: Secondary education

If the group of educands has no distinguishing quality the educational community is entered under the term for the educational level, e.g.

SECONDARY EDUCATION: STUDENTS

The convention has been employed in the Specimen Index that all the words specifying a social group are always given in capitals. This enables a distinction to be made between headings which deal with social groups and those which deal with purely educational questions. Thus we should enter

CANADIANS: Secondary education: Administration

CANADIANS: SECONDARY EDUCATION,. Students

This method of entry emphasises the social group aspect of educational questions wherever it is present. Moreover when the group is defined at any point by an educational level, other questions, peripheral for the sociologist, but relating to the corresponding education system are brought into proximity, as in the case above where CANADIANS: SECONDARY EDUCATION,. STUDENTS - a social group - is brought into proximity with CANADIANS: Secondary Education: Administration.

One further question of general importance arose in course of preparing the Specimen Index. This concerns the role of concepts meaning a place or location. In analysing the literature of physical science and technology, place concepts are encountered which simply indicate location in space. In the literature of sociology, place concepts indicate both location in space and the identity of a particular Social group. Thus item 634 deals with a question of the bilingual teaching of students who already speak Swahili and English. Clearly bilingualism in these two languages must form part or the specification of this Social group. The teaching in this case occurs in Kenya, but the underlying point of view in the Specimen Index is that in such a case Kenya is more significant as a factor in specifying a Social group i.e. Kenyans, than merely as a location. So the group is finally specified as KENYANS: SCHOOLS,. STUDENTS, BILINGUAL, SWAHILI - ENGLISH.

The generalised concept structures in evidence in the Specimen Index are the following:-

(1) Position of main elements in subject headings (Facet order)
(a) Social group (b) Action or Property (c) Active social group where (a) is passive.

(2) Specification of facet (a) or facet (c) Social group.

The properties by which a social group is specified are cited in the following order:-

2.1 Nationality

2.2 Type of educational community to which attached

2.3 Race

2.4 Language

2.5 Social class

2.6 Sex

2.7 Age

[A question worthy of consideration is whether properties 2.3 to 2.7 might take precedence over 2.2 where the educational community is designed specially for a particular social group. For example. A university for women might be given as

WOMEN: UNIVERSITIES (2.6 2.2)

while women at (mixed) universities would be given as

UNIVERSITIES,. WOMEN (2.2 2.6)

Such a distinction is occasionally in evidence in the Specimen Index. In this connection it should be borne in mind that many educational community items invariably imply particular ages and occasionally particular classes. The mechanism needs using with great care.]

The order of elements in subject headings is usually the reverse of the word order in English natural language using nouns and prepositions as far as possible, e.g.

GERMANS: SECONDARY EDUCATION,. STUDENTS; Achievement: Negative correlation with intellect

would be stated in English as

"Negative correlation *of* intellect *with* achievement *by* students in the secondary education system for Germans".

The approximate English equivalents of the punctuation operators when read in reverse heading order are:-

,	no equivalent - a type specifying operator
:	of (objective genitive) for
;	by (or a subject + verbal noun phrase, if the action precedes ; in the heading)
.,	at (the reverse of ,.)
(	of (possessive) (the reverse of .)
]	taught to

The subject heading scheme presented in the Specimen Index is the closest possible approximation to that used in *British Technology Index*, having due regard to such fundamental differences as exist between the structure of the concepts of technology and those of the social sciences. The scheme as offered may well be the best possible for some purposes but not for all. It is therefore necessary to consider the ways in which, if need be, it could be modified, without sacrifice of the essential features from which its usefulness mainly springs.

The possible modifications may be considered as amending the System at one of three possible levels.

At the first level of modification we leave the pre-ordinated subject heading structure intact, but make changes within that structure of the disposition of the descriptive matter giving the titles and other details of the individual items of literature. For reasons of economy of space the presented structure enters these descriptive details under one subject heading embodying a certain preferred order of elements, and cross refers to this subject heading from various other combinations of part combinations of these elements. Access from all significant elements is thus provided, though an enquirer who begins with any element other than that employed as leading word on the entry heading, gains access by a two stage process. It is possible to provide single step access by deleting the second half of all inversion references and repeating under the first half the appropriate descriptive details which the Specimen Index gives under one arrangement of the heading elements only. We should thus have an Index in which repeated direct entries of the material were substituted for single entry with supporting cross-references. The changes in the existing *British Technology Index* computer program needed to effect this would be trivial though it is likely that computer processing time involving extensive copying operations might be more than required under the existing system.

Conversely we could completely separate the descriptive details from the subject heading structure. There would be good reasons for this if it were desirable to arrange the record of the actual literature items in some order other than alphabetical subject. This need would arise, for instance, if the system were used as the subject index to a set of abstracts already published. As in the case previously considered the distinction between a subject heading (attached to descriptive details) and a cross-reference would disappear. Both subject headings and the first halves of former inversion cross-references would simply carry an "address" (such as an abstract serial number, or the classification symbol if the descriptive records are arranged in systematic order).

For example, entry 7 in the Specimen Index would transform to

AMERICANS. NEGROES. BOYS: Examining; Teachers, White; Approval - Disapproval 643

NEGROES (AMERICANS 643

BOYS (NEGROES (AMERICANS 643

EXAMINING: Boys (Negroes (Americans 643

TEACHERS, White; Examining: Boys (Negroes (Americans 643

WHITE TEACHERS; Examining: Boys (Negroes (American 643

APPROVAL - DISAPPROVAL; White teachers; Examining: Boys (Negroes (Americans 643

DISAPPROVAL - APPROVAL; White teachers; Examining: Boys (Negroes (Americans 643

An unavoidable corollary of structures of this form is, of course, that *all* access requires *two* searching steps by the enquirer.

The second level of possible modification of the system comprises changes of facet or element order in subject headings. In the Specimen Index the Social group is always cited before the action which the group carries on or undergoes. It would be possible to rearrange this so that the action term occupies the leading position.

This facility for modification can be generalised more fully. It is possible to reverse completely the order elements in subject headings, so that (e.g.) the subject heading

CHILDREN, 4 YEAR OLD; Achievement

is changed to

ACHIEVEMENT; 4 Year old children.

To generalise this: If in the present Specimen Index we have the subject heading

A -- : B -- : C -- followed by descriptive material

Td

it is possible to reverse the whole system to give

C -- : B -- : A --

Td

In other words subject headings and cross-references to complete headings may be made to change roles without breaking up the system.

It is important to note here that both the order of elements used in subject headings and that of the reversed subject heading represent a logically articulated string of ideas, and analogues of each may be found in natural languages which rely critically on word order to establish syntax. If however it were decided to change the facet order of subject headings by bringing to the entry word position a word occupying an internal position in the original facet order, then, because of the resultant break in logical articulation, a deeper modification of the system would be required. The essence of this would be that the logical break itself would need to be represented by a special relational operator. There would, because of the logical break, be increased difficulty in interpreting the subject headings and cross-references into their natural language equivalents.

Modification for Co-ordinate Indexing

The separate elements found in headings in the Specimen Index can be used as Feature Headings in post-co-ordinate indexes, provided that certain of them are eliminated as likely to be unsought by any enquiry. An identical selection has in fact been made in the Specimen Index in deciding which of the elements are unsought and so do not need to initiate a cross-reference. The term 'students' is such an unsought term. It is used frequently in headings, where its use as an internal element is often obligatory. Nevertheless the concept 'student' is so pervasive in literature on education that it seems improbable that any user will access the system by this term. It is therefore omitted, equally as a leading term in the cross-referencing in the pre-co-ordinated system and as a Feature Heading in a post-co-ordinated system.

Because the same concepts denoting social groups of various kinds appear in both active and passive roles in the literature of education and of sociology, the result of using simple terms as Feature Headings is likely to be a system which gives low precision retrieval. Improvement in precision would be achieved by the use in Feature Headings of the two Role Indicators 'Active' and 'Passive'.

It is tempting to ask whether the whole relational apparatus used in the pre-co-ordinated index could not be used post-co-ordinately to form a more sophisticated role indication system. Relational operators could of course be assigned Feature Cards in the same manner as concept elements, but in this form it is highly unlikely that users would ever search on them. The essence of a post-co-ordinate system is that it eliminates from the input task the whole question of the position that a concept element occupies in the statement of a composite subject. On the other hand we have seen in the pre-co-ordinated index that the position of an element in an articulated concept string is an indispensible factor in relationship indication.

The conclusion is inevitable that while the first and second phases of intellectual work mentioned at the beginning of this study (summarisation of concept, and vocabulary control) are common to the input work of both pre-co-ordinated and post-co-ordinated indexes, the pure form of the latter has no place for the third and fourth phases (syntactic and hierarchical organisation) except on the output side in the form of maps of term hierarchies and categories designed to help the user.

Compromise forms of post-co-ordination in which partially pre-co-ordinated terms are assigned to Feature Cards are, of course, readily conceivable. It would be possible to form an inverted post-co-ordination file of such pairs as

ACHIEVEMENT; AGGRESSIVE STUDENTS

AGGRESSIVE STUDENTS,. JUNIOR SCHOOLS

CHILDREN: CONSTRAINTS

at the cost of a considerable reduction in the simplicity of original input and of search manipulation which is the main advantage of co-ordinate indexes.

CONCLUSION

An outline has been given of the role of subject relationships in forming the base for both manual and computer versions of the concept indexing system used by British Technology Index.

Part II gives the minimum modifications to the British Technology Index system which it would require to handle satisfactorily the literature of educational Sociology. A Specimen Index of 108 actual documents in the field of educational Sociology is included, and it is emphasised that various rearrangements departing from the Specimen Index are possible without impairing the integrity of the underlying system.

III SPECIMEN INDEX

to Abstracts 568-667, Sociology of Education Abstracts, 1968, 4 (4), demonstrating the possible application of a computer manipulable relation-based index structure, similar to that used in British Technology Index, to material in the field of educational sociology.

4 YEAR OLD CHILDREN. See CHILDREN, 4 YEAR OLD
8-10 YEAR OLD BOYS. See BOYS, 8-10 YEAR OLD
15-17 YEAR OLD ADOLESCENTS (MIDDLE CLASSES. See MIDDLE CLASSES: Adolescents, 15-17 year old
16-18 YEAR OLD STUDENTS., Secondary education: Irish. See IRISH: SECONDARY EDUCATION,. STUDENTS, 16-18 YEAR OLD
ABILITY
Related Headings:
LANGUAGE ABILITY
ACHIEVEMENT
Related Headings:
UNDER ACHIEVEMENT
ACHIEVEMENT; 4 year old children. See CHILDREN, 4 YEAR OLD; Achievement
ACHIEVEMENT; 8-10 year old boys. See BOYS, 8-10 YEAR OLD; Achievement
ACHIEVEMENT; Aggressive students., Junior schools. See JUNIOR SCHOOLS,.Students, Aggressive; Achievement
ACHIEVEMENT; Bilingual teaching: Swahili-English bilingual students., Schools: Kenyans. See KENYANS: SCHOOLS,.STUDENTS, BILINGUAL, SWAHILI-ENGLISH: Teaching, Bilingual; Achievement
ACHIEVEMENT., Common Entrance Examination: Correlation with Achievement., General Certificate of Education. See GENERAL CERTIFICATE OF EDUCATION,.Achievement : Correlation with Common Entrance Examination achievement
ACHIEVEMENT., General Certificate of Education. See GENERAL CERTIFICATE OF EDUCATION, Achievement
ACHIEVEMENT; Students.,Schools. See SCHOOLS,.STUDENTS; Achievement
ACHIEVEMENT; Students: Secondary education: Germans. See GERMANS: SECONDARY EDUCATION., STUDENTS; Achievement
ACHIEVEMENT; Students., Training Colleges: Teachers. See TEACHERS: TRAINING COLLEGES,. STUDENTS: Achievement
ACHIEVEMENT; Withdrawn students., Junior Schools. See JUNIOR SCHOOLS,. Students, Withdrawn: Achievement
ADMINISTRATION
Related Headings:
DECENTRALISATION
ADMINISTRATION: Education. See EDUCATION: Administration
ADMINISTRATION: Schools. See SCHOOLS: Administration
ADOLESCENTS
Related Headings:
AMERICANS: UNIVERSITIES,.Students
ADOLESCENTS, 15-17 year old (Middle classes. See MIDDLE CLASSES. Adolescents, 15-17 year old

1 ADOLESCENTS; Socialisation
Tenaringer og tilpasningsvansker i hjem, skole og samfunn. C. Hambro. Norsk Pedagogisk Tidskrift, 52 (6-7) 1968 p.225-42
(Abstr. 638)

ADULT EDUCATION. See FURTHER EDUCATION
AFRICANS
Related Headings:
KENYANS

2 AFRICANS: Education
Education and scientific and technical training in Africa. Unesco Chronicle, 14 (6) 1968 p.225-232.
(Abstr. 612)

3 AFRICANS: Education: Research
Some research problems in African education. J. Social Issues, 24 (2) 1968 p.161-75.
(Abstr. 663)

AGGRESSIVE STUDENTS., Junior schools. See JUNIOR SCHOOLS,.Students, Aggressive
AGRICULTURAL COMMUNITIES,.Social integration;Teachers. See TEACHERS; Social integration., Agricultural communities
AIMS:Education. See EDUCATION:Aims
AIMS;Students.,Further education. See FURTHER EDUCATION,.Students;Aims
AMERICANS. See also UNITED STATES
AMERICANS:Agricultural communities,.Social integration;Teachers. See TEACHERS; Social integration.,Agricultural communities:Americans

4 AMERICANS.CHILDREN/Sex education;Agents:Opinion;Mothers
Maternal preference of socialisation agent for sex education.
C.L. Harter & V.W. Parrish. J. of Marriage and the Family, 30 (3) 1968. p.418-26.
(Abstr. 640)

5 AMERICANS:Education:Participation;Community:Effect of decentralisation
Community control of education. Proc. Acad. of Political Science, 29 (1) 1968. p.60-71.
(Abstr. 629)

6 AMERICANS:HIGHER EDUCATION,.Teachers:Opinion;Students
Student image of the teacher. D.T. Hall & E.H. Schein. J. Applied Behavioural Science, 3 (3) 1967. p.305-337.
(Abstr. 637)

7 AMERICANS.NEGROES.BOYS:Examining;Teachers,White;Approval-Disapproval
Effects of race of tester, approval-disapproval, and need on negro children's learning.
T. Henchy & H. Allen. J. Personality and Social Psychology, 8 (1) 1968. p.38-42.
(Abstr. 643)

8 AMERICANS:Nursery schools,.Children:Constraints
Varieties of constraint in a nursery school. Young Children, 23 (6) 1968 p.358-67.
(Abstr. 661)

9 AMERICANS:PRIMARY EDUCATION;HEAD TEACHERS
Changing world of the principal. National Elementary Principal, 47 (5&6) 1968
(Abstr. 598)

10 AMERICANS:SCHOOLS.GOVERNING BODIES;Policies:Opinion;Social classes
Social class and school board expectations. F.D. Carver. Urban Education, 3 (3) 1968 p.14[illegible]-55
(Abstr. 596)

11 AMERICANS:SCHOOLS,.STUDENTS/Social questions
Morals, ideology and the schools. Educational Theory, 17 (4) 1967 p.271-88.
(Abstr. 635)

12 AMERICANS:UNIVERSITIES,.STUDENTS,FIRST BORN:Incidence
Birth order and college attendance in a cross-cultural setting. R.L. Greene & J.R. Clark.
J. Social Psychology, 7s (2) 1968 p.289-90.
(Abstr. 636)

13 AMERICANS:UNIVERSITIES,STUDENTS,FIRST YEAR:Unsupervised study programmes
College innovation and the decline of the West. J. Applied Behavioural Science, 3 (3) 1967 p.339-45.
(Abstr. 665)

14 AMERICANS:UNIVERSITIES,.STUDENTS:Social groups:Formation
Selected problems of adolescence: with special emphasis on group formation. H. Deutsch.
London, Hogarth P. & Inst. of Psychoanalysis, 1968
(Abstr. 607)

APPROVAL-DISAPPROVAL;White teachers;Examining:Boys(Negroes(Americans.
See AMERICANS.NEGROES.BOYS:Examining;Teachers,White;Approval-Disapproval

15 ASIANS.Society
Application of science and technology to the development of Asia.
Unesco Chronicle, 14 (7&8) 1968 p.267-72.
(Abstr. 572)

ATTENDANCE;Students.,Further education. See FURTHER EDUCATION,.STUDENTS;Attendance
ATTENDANCE;Students.,Liberal arts/Universities. See UNIVERSITIES/Liberal arts,.Students; Attendance
ATTITUDES;Students.,Higher education. See HIGHER EDUCATION,.Students;Attitudes

16 AUSTRALIANS:Teachers;Job satisfaction
Need structure of teachers varying in experience and job satisfaction. J. Educational Administration, 6 (1) 1968 p.41-51.
(Abstr. 610)

17 AUSTRALIANS:UNIVERSITIES,.STUDENTS;Religious activities.Change
Changes in religious activity among students at an Australian university. K.C. Dempsey & J. Pandey. Australian J. Higher Education, 3 (1) 1967 p.78-82.
(Abstr. 606)

BEHAVIOUR:Modification
Related Headings:
CONDITIONING
BIBLIOGRAPHIES:Higher education/Attitudes;Students.,Higher education. See HIGHER EDUCATION,.STUDENTS:Attitudes/Higher education:Bibliographies
BIBLIOGRAPHIES:Higher education planning. See HIGHER EDUCATION:Planning:Bibliographies
BIBLIOGRAPHIES:Politics/Attitudes;Students.,Higher education. See HIGHER EDUCATION,.STUDENTS;Attitudes/Politics:Bibliographies
BILINGUAL STUDENTS.,Schools:Kenyans. See KENYANS:SCHOOLS,.Students,Bilingual
BILINGUAL TEACHING:Greenlanders. See GREENLANDERS:Teaching,Bilingual
BILINGUAL TEACHING:Swahili-English bilingual students.,Schools:Kenyans. See KENYANS:SCHOOLS,.STUDENTS,BILINGUAL,Swahili-English:Teaching,Bilingual
BIRTH ORDER GROUPS
Related Headings:
FIRST BORN

18 BIRTH ORDER GROUPS;Ethical sense
Birth order and responsibility. I.D. Harris & K.I. Howard. J. Marriage and the Family, 30 (3) 1968 p.427-32.
(Abstr. 639)

19 BOYS,8-10 YEAR OL[illegible]chievement:Effect of Achievement:Attitude(Mothers
Longitudinal study of the origin of achievement strivings. S.C. Feld. J. Personality & Social Psychology, 7 (4) 1967
(Abstr. 619)

BOYS(Negroes(Americans. SEE AMERICANS.NEGROES.BOYS
BRITONS. See also GREAT BRITAIN
BRITONS:Educational measurements
Related Headings:
COMMON ENTRANCE EXAMINATION
GENERAL CERTIFICATE OF EDUCATION

20 BRITONS:Further education
Aims and organisation of further education. D.F. Bratchell. Oxford, Pergamon Press, 1968
(Abstr. 587)

BRITONS:Primary education
Related Headings:
JUNIOR SCHOOLS

21 BRITONS:Schools:Leavers;Migration:Effect of Educational grades
Where did the school leavers go? J.W. House, A.D. Thomas & K.G. Willis. University of Newcastle-upon-Tyne, Geography Dept,. 1968
(Abstr. 654)

BRITONS:Secondary education
Related Headings:
COMPREHENSIVE SCHOOLS
GRAMMAR SCHOOLS
PUBLIC SCHOOLS
TECHNICAL COLLEGES

22 BRITONS:Technical education
Education for industry, ed. by D. Hutchings. London, Longmans, Green, 1968.
(Abstr. 657)

23 CANADIANS:Education
Canada: report on educational developments in 1967-8. Canadian Education Association. Canadian Education and Research Digest, 8 (2) 1968 p.175-87.
(Abstr. 593)

24 CANADIANS:SCHOOLS.GOVERNING BODIES;Verbal interaction patterns
Adaptation of the Bellack classroom interaction analysis technique to the school board meeting. E.A. Holdaway & D.A. Mackay. Canadian Education and Research Digest, 8 (2) 1968 p.141-7.
(Abstr. 652)

25 CANADIANS:SCHOOLS,.SUPERINTENDENTS:Relations with governing bodies
Decision-making and the politics of education. E.D. Hodgson. Canadian Education and Research Digest, 8 (2) 1968 p.150-62. (Abstr. 650)

CAREERS:Counselling. See VOCATIONAL GUIDANCE

26 CAREERS.Developmental psychology
Life-stage vocational development system. D.B. Hershenson. J. Counselling Psychology, 15 (1) 1968 p.23-30. (Abstr. 644)

CAREERS.,Organisations. See ORGANISATIONS,.Careers

CEYLON. See SINGHALESE

CHILD-PARENT RELATIONS:Correlation with achievement;Women students.,Universities. See UNIVERSITIES,.STUDENTS,WOMEN;Achievement:Correlation with child-parent relations

CHILDREN
Related Headings:
BOYS

27 CHILDREN,4 YEAR OLD;Achievement:Correlation with achievement:Expectation;Mothers
Achievement motive of 4-year-olds and maternal achievement expectancies. J. Experimental Education, 36 (2) 1968 p.14-23. (Abstr. 592)

CHILDREN(Americans. See AMERICANS.CHILDREN

28 CHILDREN;Delinquency:Designation;Juridical bodies
Social organisation of juvenile justice. A.V. Cicourel. London & New York, John Wiley, 1968. (Abstr. 600)

CHILDREN:Education
Related Headings:
JUNIOR SCHOOLS
NURSERY SCHOOLS
PRIMARY EDUCATION
SCHOOLS
SECONDARY EDUCATION

CHILDREN(LOW INCOME GROUPS. See LOW INCOME GROUPS. Children

CHILDREN.,Nursey schools:Americans. See AMERICANS: NURSERY SCHOOLS., CHILDREN

CLASSES,Social. See SOCIAL CLASSES

CLASSES,Social;Opinion:Policies:American governing bodies(Schools. See SCHOOLS.GOVERNING BODIES,American;Policies:Opinion;Social classes

COMMON ENTRANCE EXAMINATION,.Achievement:Correlation with Achievement.,General Certificate of Education. See GENERAL CERTIFICATE OF EDUCATION,.Achievement:Correlation with Common Entrance Examination achievement

COMMUNICATING:Teaching;Students.,Training colleges:Teachers. See TEACHERS:TRAINING COLLEGES,.STUDENTS;Teaching:Communicating

COMMUNITY PARTICIPATION:Education:Americans. See AMERICANS:Education:Participation; Community

29 COMPREHENSIVE SCHOOLS
Risinghill: death of a comprehensive school. L. Berg. Harmondsworth, Penguin Books. 1968 (Abstr. 577)

30 COMPREHENSIVE SCHOOLS:Development
Long way from comprehensive. A. Corbett. New Society, 12 (302) 1968 p.51-2. (Abstr. 602)

CONDITIONING
Related Headings:
REINFORCEMENT

CONSTRAINTS:Children.,Nursery schools:Americans. See AMERICANS:NURSERY SCHOOLS,.CHILDREN

31 CONTINUOUS LEARNING
Continuous learning through the educative community: an exploration of the family-educational, the sequential-unit, and the complementary functional systems. P.L. Essert & R.F. Spence. Adult Education J. 18 (4) 1968 p.260-71. (Abstr. 617)

COOPERATION:Research. See RESEARCH:Cooperation

COUNSELLING
Related Headings:
VOCATIONAL GUIDANCE

CULTURE;Danish-Eskimo bilingual teaching:Greenlanders. See GREENLANDERS:Teaching,Bilingual, Danish-Eskimo;Culture

CUMULATIVE LEARNING. See LEARNING,Cumulative
DANISH-ESKIMO BILINGUAL TEACHING:Greenlanders. See GREENLANDERS:Teaching,Bilingual, Danish-Eskimo
DECENTRALISATION:Effect on community participation:Education:Americans. See AMERICANS: Education:Participation;Community:Effect of decentralisation
DELINQUENCY;15-17 year old adolescents(MIDDLE CLASSES. See MIDDLE CLASSES. ADOLESCENTS,15-17 year old;Delinquency
DELINQUENCY;Children. See CHILDREN;Delinquency
DELINQUENCY;New Zealanders. See NEW ZEALANDERS:Delinquency
DEVELOPMENT:Intellect. See INTELLECT:Development
DEVELOPMENT:Psychology
  Related Headings:
    MATURATION
DEVELOPMENT.Psychology(Careers. See CAREERS.Developmental psychology

32 DEWEY, John. Educational theory:Adaptation;Progressive groups..United States
  Ideologising of theories: John Dewey's educational theory, a case in point. G. Eastman. Educational theory, 17 (2) 1967 p.103-119.
  (Abstr. 611)

DISAPPROVAL-APPROVAL;White teachers;Examining.Boys(Negroes(Americans. See AMERICANS.NEGROES. BOYS:Examining;Teachers,White;Approval-Disapproval
DISCONTINUITY(Culture;Danish-Eskimo bilingual teaching:Greenlanders. See GREENLANDERS: Teaching,Bilingual,Danish-Eskimo;Culture.Discontinuity
DOMINATION;Individuals(Students.,Schools. See SCHOOLS.,STUDENTS.Individuals;Domination: Modification:Reinforcement
EARLY LEAVERS:Schools:Norwegians. See NORWEGIANS:SCHOOLS:EARLY LEAVERS
ECONOMICS
  Related Headings:
    LABOUR

33 EDUCATION:Administration.Sociology
  Educational administration as a social process: theory, research, practice. J.W. Getzels, J.M. Lipham & R.F. Campbell. London & New York, Harper & Row. 1968
  (Abstr. 628)

EDUCATION:Africans. See AFRICANS:Education

34 EDUCATION:Aims
  Freedom and the school. C.W. Hobart. Canadian Education and Research Digest, 8 (3) 1968 p.269-82
  (Abstr. 647)

35 Learning Society. R.M. Hutchins. London, Pall Mall, 1968
  (Abstr. 658)

EDUCATION:Americans. See AMERICANS:Education
EDUCATION:Canadians. See CANADIANS:Education
EDUCATION:Children. See CHILDREN:Education
EDUCATION:Effect on Manpower(Indians. See INDIANS.Manpower:Effect of education
EDUCATION:French people. See FRENCH PEOPLE:Education
EDUCATION.Grades:Effect on Migration;Leavers:Schools:Britons. See BRITONS:Schools:Leavers; Migration:Effect of educational grades
EDUCATION.Grades:Effect on Migration;Leavers:Technical colleges. See TECHNICAL COLLEGES: Leavers;Migration:Effect of Educational Grades
EDUCATION:Low income groups. See LOW INCOME GROUPS:Education
EDUCATION:Measurements
  Related Headings:
    EXAMINING
EDUCATION:Minorities. See MINORITIES:Education

36 EDUCATION:Planning
  Educational planning: achievements and prospects. Unesco Chronicle, 14 (7-8) 1968 p.273-8
  (Abstr. 613)

37 EDUCATION:Research
  Research in education. J.L. Hayman. Ohio, Charles E. Merrill, 1968
  (Abstr. 641)

38 EDUCATION:Research..Sweden
  Educational research in Sweden: some areas of current and potential development. A. Bjerstedt. International Rev. of Education, 14 (3) 1968 p.259-76.
  (Abstr. 580)

39 EDUCATION:Seeking:Migration
Emigration, education and manpower. J.A. Jackson. European Teacher, 6 (2) 1968 p.5-9
(Abstr. 660)

40 EDUCATION.Sociology
Sociology of education. O. Banks. London, Batsford 1968
(Abstr. 575)

EDUCATION:Swedes. See SWEDES:Education
EDUCATION:Teachers. See TEACHERS:Education
EDUCATION.Theory(Dewey, John. See DEWEY, John. Educational theory
EDUCATIONAL STATUS:Correlation with attendance:Students.,Further education. See FURTHER EDUCATION,.STUDENTS;Attendance:Correlation with educational status
ENGLISH/Students.,Fifth forms(Grammar schools. See GRAMMAR SCHOOLS.FIFTH FORMS,.STUDENTS /English
ENGLISH-SWAHILI BILINGUAL STUDENTS.,Schools:Kenyans. See KENYANS:Schools,.Students, Bilingual,Swahili-English
ESKIMO-DANISH BILINGUAL TEACHING:Greenlanders. See GREENLANDERS,Teaching,Bilingual, Danish-Eskimo
ETHICAL SENSE;Birth order groups. See BIRTH ORDER GROUPS:Ethical sense
EUROPEAN CULTURE/Teaching: 16-18 year old students.,Secondary education:Irish. See IRISH: SECONDARY EDUCATION,.STUDENTS, 16-18 year old:Teaching/European culture
EXAMINING:Boys(Negroes(Americans. See AMERICANS.NEGROES.BOYS:Examining
EXPERIMENTS
Related Headings:
FACTOR ANALYSIS
QUESTIONNAIRES
FACTOR ANALYSIS:Individualism studies:Intellect:Negative correlation with achievement: Students.,Secondary education:Germans. See GERMANS:SECONDARY EDUCATION,.STUDENTS; Achievement:Negative correlation with intellect.Individualism:Studies;Factor analysis
FARMING COMMUNITIES. See AGRICULTURAL COMMUNITIES
FIFTH FORMS(Grammar schools. See GRAMMAR SCHOOLS.FIFTH FORMS
FIRST BORN STUDENTS.,Schools. See SCHOOLS,.STUDENTS,FIRST BORN
FIRST BORN STUDENTS.,Universities:Americans. See AMERICANS:Universities,.Students,First born
FIRST YEAR STUDENTS.,Universities:Americans. See AMERICANS:UNIVERSITIES,.STUDENTS,FIRST YEAR:Unsupervised study programmes
FRANCE. See FRENCH
FRENCH JOURNAL ARTICLES/Further education. See FURTHER EDUCATION/Journal articles, French
FRENCH JOURNAL ARTICLES/Secondary education. See SECONDARY EDUCATION/Journal articles,French

41 FRENCH PEOPLE:Education
Tomorrow's education: the French experience. J. Capelle. Oxford, Pergamon Press, 1967
(Abstr. 594)

FURTHER EDUCATION
Related Headings:
CONTINUOUS LEARNING
FURTHER EDUCATION:Britons. See BRITONS:Further education

42 FURTHER EDUCATION/Journal articles,French
Enseignements techniques, 2 1968
(Abstr. 614)

43 FURTHER EDUCATION,.STUDENTS;Aims
Blessed wantfulness of the common people. J. Borley. Technical J., 6 (4) 1968 p.10-11.
(Abstr. 585)

44 FURTHER EDUCATION,.STUDENTS;Attendance:Correlation with educational status
Differential participation patterns of adults of low and high educational attainment. M. Douglah & G. Moss. Adult Education J. 18 (4) 1968 p.247-59.
(Abstr. 609)

G.C.E. See GENERAL CERTIFICATE OF EDUCATION

45 GENERAL CERTIFICATE OF EDUCATION,.Achievement:Correlation with Common Entrance Examination achievement
Intelligence, Common Entrance, and General Certificate of Education. S.W. Hockey. Br. J. Educational Psychology, 38 (2) 1968 p.140-7.
(Abstr. 649)

46 GERMANS:SECONDARY EDUCATION,.STUDENTS;Achievement:Negative correlation with intellect.
Individualism:Studies:Factor analysis
Faktorenanalytische Untersuchungen über die Erfolgsbedingungen an verschiedenen Gymnasien.
R. Burger. Pädagogische Rundschau, 22 (3-4) 1968 p.134-49.
(Abstr. 589)

GOVERNING BODIES:Relations with Superintendents.,Schools:Canadians. See CANADIANS:Schools,.
Superintendents:Relations with governing bodies
GOVERNING BODIES(Schools:Americans. See AMERICANS:SCHOOLS.GOVERNING BODIES
GOVERNING BODIES(Schools:Canadians. See CANADIANS:SCHOOLS.GOVERNING BODIES
GOVERNING BODIES(Universities:Relations with students.,Universities. See UNIVERSITIES,.
STUDENTS:Relations with university governing bodies
GOVERNMENTS:Relations with universities. See UNIVERSITIES: Relations with governments

47 GRAMMAR SCHOOLS. FIFTH FORMS,.STUDENTS/English;Under achievement:Correlation with parent
qualities
Home background -- its effect on achievement in English. J.R. Birnie & E.A. Hewitt.
Educational Rev., 20 1968 p.156-63
(Abstr. 578)

GREAT BRITAIN
See
IMMIGRANTS:SCHOOLS,.STUDENTS..Great Britain
GREAT BRITAIN. See also BRITONS

48 GREENLANDERS:Teaching,Bilingual,Danish-Eskimo;Culture.Discontinuity
Influence of the school on acculturation with special reference to Greenland.
C.W. Hobart. J. Educational Thought, 2 (2) 1968 p.97-116
(Abstr. 648)

HEAD TEACHERS.Opinion:Aims(Teaching. See TEACHING.Aims:Opinion(Head teachers
HEAD TEACHERS;Primary education:Americans. See AMERICANS:PRIMARY EDUCATION;HEAD TEACHERS
HIGHER EDUCATION
Related Headings:
UNIVERSITIES
HIGHER EDUCATION:Americans. See AMERICANS:HIGHER EDUCATION
HIGHER EDUCATION/Attitudes;Students.,Higher education. See HIGHER EDUCATION,.STUDENTS;
Attitudes/Higher education

49 HIGHER EDUCATION:Planning:Bibliographies
Economics of higher education and educational planning -- a bibliography. Socio-Economic
Planning Sciences, 2 (1) 1968 p.25-101
(Abstr. 655)

50 HIGHER EDUCATION,.Students;Attitudes/Higher education:Bibliographies
Select bibliography on students politics and higher education.
P. Altbach. Cambridge (Mass.) Center for International Affairs, Harvard University, 1967.
(Abstr. 570)

51 HIGHER EDUCATION,.Students;Attitudes/Politics:Bibliographies
Select bibliography on students politics and higher education. P. Altbach. Cambridge
(Mass.) Center for International Affairs, Harvard University, 1967.
(Abstr. 570)

52 IMMIGRANTS:SCHOOLS,.STUDENTS..Great Britain
Immigrant education. A. Corbett. New Society, 12 (311) 1968 p.361-2
(Abstr. 603)

53 INDIANS.Manpower:Effect of education
Manpower and economic development in India, 1961-8. T. Burgess, R. Layard & P. Pant.
Edinburgh, Oliver & Boyd, 1968
(Abstr. 590)

INDIVIDUALISM(Intellect:Negative correlation with achievement;Students., Secondary
education:Germans. See GERMANS:SECONDARY EDUCATION,.Students;Achievement:Negative
correlation with intellect.Individualism
INDIVIDUALS(Students.,Schools. See SCHOOLS,.STUDENTS

54 INNOVATION. Sociology
General theory of innovativeness. H.D. Hemphill. Alberta J. of Educational Research,
14 (2) 1968 p.101-14
(Abstr. 642)

INTEGRATION,Social;Teachers. See TEACHERS;Social integration

55 INTELLECT:Development
Organisation of abilities and the development of intelligence. J.L. Horn.
Psychological Rev., 75 (3) 1968 p.242-59
(Abstr. 653)

INTELLECT:Development
Related Headings:
LEARNING,Cumulative

INTELLECT:Negative correlation with achievement;Students.,Secondary education:Germans. See GERMANS:SECONDARY EDUCATION,.STUDENTS;Achievement:Negative correlation with intellect

56 IRISH:SECONDARY EDUCATION,.STUDENTS, 16-18 YEAR OLD:Teaching/European culture
European studies in Irish schools. H.R. Cathcart. European Teacher, 6 (3) 1968 p.5-9.
(Abstr. 597)

JOB SATISFACTION;Teachers:Australians. See AUSTRALIANS:Teachers;Job satisfaction

JOURNAL ARTICLES,French/Further education. See FURTHER EDUCATION/Journal articles, French

JOURNAL ARTICLES,French/Secondary education. See SECONDARY EDUCATION/Journal articles,French

57 JUNIOR SCHOOLS,.STUDENTS,AGGRESSIVE;Achievement
Association of attainment and sociometric status with reported deviant behaviour in a progressive and a traditional junior school. S. Gooch & Kellmer Pringle,M.L. Br. J. Social and Clinical Psychology, 7 (3) 1968 p.184-93.
(Abstr. 632)

58 JUNIOR SCHOOLS,.STUDENTS,WITHDRAWN;Achievement
Association of attainment and sociometric status with reported deviant behaviour in a progressive and a traditional junior school. S. Gooch & Kellmer Pringle,M.L. Br. J. Social and Clinical Psychology, 7 (3) 1968 p.184-93.
(Abstr. 632)

JURIDICAL BODIES:Delinquency designation;Children. See CHILDREN;Delinquency:Designation; Juridical bodies

59 KENYANS:SCHOOLS,.STUDENTS,BILINGUAL,SWAHILI-ENGLISH:Teaching,Bilingual;Achievement
Bilingualism in the educational system of Kenya. T.P. Gorman. Comparative Education, 4 (2) 1968 p.213-20.
(Abstr. 634)

LABOUR
Related Headings:
MANPOWER

LANGUAGE ABILITY
Related Headings:
VERBAL ABILITY

LAW
Related Headings:
JURIDICAL

60 LEARNING,Cumulative
Contributions of learning to human development. R.M. Gagne. Psychological Rev., 75 (3) 1968 p.177-9.
(Abstr. 626)

LITERARY MEN;Opinion:Society. See SOCIETY:Opinion;Literary men

LOW INCOME GROUPS
Related Headings:
SLUM DWELLERS

61 LOW INCOME GROUPS.CHILDREN.Sociology
Children and poverty: some sociological and psychological perspectives. Chicago, Rand McNally, 1968
(Abstr. 631)

62 LOW INCOME GROUPS:Education.Aims
Disadvantaged: challenge to education. M.D. Fantini & G. Weinstein. New York and London, Harper & Row, 1968
(Abstr. 618)

LOWER MIDDLE CLASS STUDENTS.,Public schools. See PUBLIC SCHOOLS,.STUDENTS,Lower middle class

MANPOWER(Indians. See INDIANS.Manpower

MATURATION;Students.,Universities. See UNIVERSITIES,.STUDENTS;Maturation

63 MIDDLE CLASSES.ADOLESCENTS,15-17 year old;Delinquency:Correlation with delinquency: Acceptance;Parents
Parental attitudes of sanction in middle class adolescent male delinquency. C.R. Gallenkamp & J.F. Rychlak. J. Social Psychology, 75 (2) 1968 p.255-60.
(Abstr. 627)

MIGRATION
Related Headings:
IMMIGRANTS

MIGRATION:Education seeking. See EDUCATION:Seeking:Migration
MIGRATION;Leavers:Schools:Britons. See BRITONS:Schools:Leavers;Migration
MIGRATION;Leavers:Technical colleges. See TECHNICAL COLLEGES:Leavers;Migration
MIGRATION;Professional classes. See PROFESSIONAL CLASSES;Migration
MIGRATION;Technicians. See TECHNICIANS;Migration

64 MINORITIES:Education.Aims
Disadvantaged: challenge to education. D.M. Fantani & G. Weinstein. New York and London, Harper & Row, 1968
(Abstr. 618)

MOBILITY.,Social classes. See SOCIAL CLASSES,.Mobility
MOBILITY.,Social classes;First born students,.Schools. See SCHOOLS,.STUDENTS,FIRST BORN; Social mobility
MOTHERS.Attitude:Achievement:Effect on Achievement;8-10 year old boys. See BOYS,8-10 year old;Achievement:Effect of Achievement:Attitude(Mothers
MOTHERS;Expectation:Achievement:Correlation with achievement:4 year old children. See CHILDREN,4 YEAR OLD;Achievement:Correlation with achievement:Expectation;Mothers
MOTHERS;Opinion:Sex education agents/ Children(Americans. See AMERICANS.CHILDREN/Sex education;Agents:Opinion;Mothers
MOTIVATION:Correlation with achievement;Students.,Schools. See SCHOOLS,.STUDENTS; Achievement:Correlation with motivation
NEGROES(Americans. See AMERICANS.NEGROES

65 NEW ZEALANDERS;Delinquency.Rate
Changing rate of delinquency in New Zealand. G.L. Arvidson & E.W. Braithwaite. New Zealand J. of Educational Studies, 3 (1) 1968 p.73-76.
(Abstr. 573)

66 New Zealand delinquency rate and the birth year effect. J. Fleming & S.W. Slater. New Zealand J. of Educational Studies, 3 (1) 1968 p.59-72.
(Abstr. 622)

67 NORWEGIANS:Schools:Early leavers
Skoletrøtte eleven. K. Jørgensen. Norsk Pedagogisk Tidskrift, 52 (6-7) p.250-2.
(Abstr. 667)

NURSERY SCHOOLS:Americans. See AMERICANS:Nursery schools
OBJECTIVES. See AIMS
OCCUPATIONAL GROUPS
Related Headings:
AGRICULTURAL GROUPS
OCCUPATIONS
Related Headings:
JOB SATISFACTION
OCCUPATIONS/Counselling. See VOCATIONAL GUIDANCE

68 ORGANISATIONS,.Careers
Organisational careers: a source book for theory. B.G. Glaser. Chicago, Aldine Publishing Co., 1968
(Abstr. 630)

69 ORGANISATIONS.Sociology
Sociology of education. O. Banks. London, Batsford 1968
(Abstr. 575)

70 Readings in organisation theory: a behavioural approach. W.A. Hill & D.M. Egan. Boston, Allyn & Bacon, 1967
(Abstr. 645)

PARENT-CHILD RELATIONS:Correlation with achievement;Women students.,Universities. See UNIVERSITIES,.STUDENTS,WOMEN;Achievement:Correlation with child-parent relations
PARENTS;Acceptance:Delinquency:Correlation with delinquency; 15-17 year old adolescents (Middle classes. See MIDDLE CLASSES.Adolescents,15-17 year old;Delinquency:Correlation with Delinquency:Acceptance;Parents
PARENTS.Qualities:Correlation with under achievement;English/Students.,Fifth forms (Grammar Schools. See GRAMMAR SCHOOLS.FIFTH FORMS,.STUDENTS/English;Under achievement: Correlation with parent qualities
PICTURE TESTS:Child-parent relations:Correlation with achievement:Women students., Universities. See UNIVERSITIES,.STUDENTS,WOMEN;Achievement:Correlation with child-parent relations:Picture tests

POLITICAL GROUPS
  Related Headings:
    PROGRESSIVE GROUPS
POLITICS
  Related Headings:
    GOVERNMENTS
POLITICS/Attitudes;Students.,Higher education. See HIGHER EDUCATION,.STUDENTS;Attitudes /Politics
POOR. See LOW INCOME GROUPS
PRIMARY EDUCATION:Americans. See AMERICANS:PRIMARY EDUCATION
PRIMARY EDUCATION:Britons. See BRITONS:Primary education

71 PROFESSIONAL CLASSES;Migration
  "Brain drains" in an integrating Europe. R. Blandy. Comparative Education Rev., 12 (2) 1968 p.180-193
  (Abstr. 582)

PROFESSIONS(Teaching. See TEACHING.Profession
PROGRESSIVE GROUPS;Adaptation:Educational theory (Dewey, John. See DEWEY, John. Educational theory:Adaptation;Progressive groups
PSYCHOLOGY
  Related Headings:
    ABILITY
    BEHAVIOUR
    DEVELOPMENT.Psychology

72 PSYCHOLOGY Experiments. Subject-Experimenter Interaction
  Social nature of sociological research: the psychological experiment as a social interaction. N. Friedman. New York and London, Basic Books, 1967
  (Abstr. 624)

73 PSYCHOLOGY:Questionnaires:Analysis;Rasch models
  Rash'ske Modeller och deres anvendelse på psykologiska data. E.B. Anderson. Nordisk Psykologi, 3 1968 p.122-38.
  (Abstr. 571)

PSYCHOLOGY:Typology
  Related Headings:
    AGGRESSIVE
    WITHDRAWN
PUBLIC SCHOOLS,.Social integration;Lower middle class students. See PUBLIC SCHOOLS,.STUDENTS,LOWER MIDDLE CLASS;Social integration
PUBLIC SCHOOLS,.Social integration;Working class students. See PUBLIC SCHOOLS,.STUDENTS,WORKING CLASS;Social integration

74 PUBLIC SCHOOLS,.STUDENTS,LOWER MIDDLE CLASS;Social integration
  Outsiders inside the public schools. J. Hipkin. New Society, 12 (303) 1968 p.79-81.
  (Abstr. 646)

75 PUBLIC SCHOOLS,.STUDENTS,WORKING CLASS;Social Integration
  Outsiders inside the public schools. J. Hipkin. New Society, 12 (303) 1968 p.79-81.
  (Abstr. 646)

QUESTIONNAIRES:Maturation;Students.,Universities. See UNIVERSITIES,.STUDENTS;Maturation:Questionnaires
QUESTIONNAIRES:Psychology. See PSYCHOLOGY:Questionnaires
RASCH MODELS:Questionnaire analysis:Psychology. See PSYCHOLOGY:Questionnaires:Analysis;Rasch models
REFERENCE GROUPS;Sociology research. See SOCIOLOGY:Research;Reference groups
REINFORCEMENT:Domination:Modification:Individuals(Students.,Schools. See SCHOOLS,.Students. Individuals;Domination:Modification:Reinforcement
RELIGIOUS ACTIVITIES;Students.,Universities:Australians. See AUSTRALIANS:UNIVERSITIES,.STUDENTS;Religious activities

76 RESEARCH:Cooperation
  Problems of research collaboration and co-operation. J.A. Fishman. J. Social Issues, 24 (2) 1968 p.235-41.
  (Abstr. 621)

RESEARCH:Education. See EDUCATION:Research
RESEARCH:Science/Universities. See UNIVERSITIES/Science:Research
ROLES
  Related Headings:
    AIMS

SCHOOL LEAVERS:Britons. See BRITONS:Schools:Leavers
SCHOOL LEAVERS,Early:Norwegians. See NORWEGIANS:SCHOOLS:EARLY LEAVERS

77 SCHOOLS:Administration
Art of the headmaster: with suggestions on school administration. F.J. Goodwin. London, Ward Lock Educational, 1968
(Abstr. 633)

SCHOOLS:Americans. See AMERICANS:Schools
SCHOOLS:Canadians. See CANADIANS:Schools
SCHOOLS:Immigrants. See IMMIGRANTS:Schools
SCHOOLS:Kenyans. See KENYANS:SCHOOLS

78 SCHOOLS,.STUDENTS;Achievement:Correlation with motivation
Academic motivation and school attainment. N.J. Entwistle. J. Educational Psychology, 38 (2) 1968 p.181-8.
(Abstr. 616)

79 SCHOOLS,.STUDENTS,FIRST BORN;Social mobility
Ordinal position and social mobility. C.N. Alexander,Jr. Sociometry, 31 (3) 1968 p.285-93.
(Abstr. 568)

80 SCHOOLS,.STUDENTS.Individuals;Domination:Modification:Reinforcement
Modification of dominant behaviour in school children. E.R. Blum & W.A. Kennedy. J. Personality and Social Psychology, 7 (3) 1967 p.275-81.
(Abstr. 583)

SCIENCE:Effect on Society(Asians. See ASIANS.Society:Effect of science
SCIENTIFIC METHOD:Effect on Society. See SOCIETY:Effect of scientific method
SECONDARY EDUCATION:Britons. See BRITONS:Secondary education
SECONDARY EDUCATION:Effect on mobility.,Social classes. See SOCIAL CLASSES,.Mobility: Effect of secondary education
SECONDARY EDUCATION:Germans. See GERMANS:Secondary education
SECONDARY EDUCATION:Irish. See IRISH:Secondary education

81 SECONDARY EDUCATION/Journal articles,French:Lists
Enseignements techniques, 2. 1968 p.79-83.
(Abstr. 614)

SEX EDUCATION/Children(Americans. See AMERICANS.CHILDREN/Sex education

82 SINGHALESE:UNIVERSITIES,.STUDENTS.Vocational unfitness
Educational dilemmas of a developing country. J.E. Jayasuriya. J. Social Sciences, 24 (2) 1968 p.199-205
(Abstr. 664)

83 SLUM DWELLERS.CHILDREN.Verbal ability
Developmental sociolinguistics: inner-city children. D.R. Entwistle. American J. of Sociology, 74 (1) 1968 p.37-49.
(Abstr. 615)

SOCIAL CLASSES
Related Headings:
LOW INCOME GROUPS
LOWER MIDDLE CLASS
MIDDLE CLASSES
WORKING CLASS

84 SOCIAL CLASSES,.Mobility:Effect of Secondary education
Gesellschaftswandel und Schulsystem in modernen Industriestaat. W. Buhl. International Rev. of Education, 14 (3) 1968 p.277-99.
(Abstr. 588)

SOCIAL CLASSES,.Mobility;First born students.,Schools. See SCHOOLS,.STUDENTS,FIRST BORN; Social mobility
SOCIAL CLASSES;Opinion:Policies;Governing bodies(Schools:Americans. See AMERICANS:SCHOOLS. GOVERNING BODIES;Policies:Opinion;Social classes
SOCIAL GROUPS
Related Headings:
BIRTH ORDER GROUPS
OCCUPATIONAL GROUPS
ORGANISATIONS
POLITICAL GROUPS
PROFESSIONAL CLASSES
SOCIAL CLASSES
TECHNICIANS

SOCIAL GROUPS:Interactions
Related Headings:
INDIVIDUALS

SOCIAL INTEGRATION;Teachers. See TEACHERS;Social integration

SOCIAL QUESTIONS/Students.,Schools:Americans. See AMERICANS:SCHOOLS,.STUDENTS/Social questions

SOCIALISATION
Related Headings:
COOPERATION

SOCIALISATION;Adolescents. See ADOLESCENTS;Socialisation

85 SOCIETY:Effect of scientific method
Cultural role of scientific behaviour. N.P. Jacobson. Educational Theory, 18 (1) 1968 p.23-31.
(Abstr. 662)

86 SOCIETY:Opinion;Literary men
Literary imagination and the study of society. R. Hoggart. Birmingham Centre for Contemporary Cultural Studies, Birmingham University, 1968
(Abstr. 651)

SOCIOLOGY (Administration:Education. See EDUCATION:Administration.Sociology

SOCIOLOGY (Children (Low income groups. See LOW INCOME GROUPS.CHILDREN.Sociology

SOCIOLOGY (Education. See EDUCATION.Sociology

SOCIOLOGY/Education:Teachers. See TEACHERS:Education/Sociology

SOCIOLOGY (Innovation. See INNOVATION.Sociology

SOCIOLOGY (Organisations. See ORGANISATIONS.Sociology

87 SOCIOLOGY:Research;Reference groups
Readings in reference group theory and research. H.H. Hyman & E. Singer. New York, Free Press;London, Collier-Macmillan, 1968
(Abstr. 659)

STATUS GROUPS. See SOCIAL CLASSES

STRATIFICATION,Social. See SOCIAL CLASSES

STUDENT OPINION:Teachers.,Higher education:Americans. See AMERICANS:Higher education,.Teachers:Opinion;Students

SUPERINTENDENTS.,Schools:Canadians. See CANADIANS:Schools,.Superintendents

SWAHILI-ENGLISH BILINGUAL STUDENTS.,Schools:Kenyans. See KENYANS:SCHOOLS,.STUDENTS, BILINGUAL,Swahili-English

SWEDEN
See
EDUCATION:Research..Sweden

88 SWEDES:Education.History,Since 1940
Educational research and educational change: the case of Sweden. T. Husen & G. Boalt. London and New York, John Wiley: Stockholm, Almquist & Wiksell, 1968
(Abstr. 656)

TEACHERS
Related Headings:
HEAD TEACHERS
SUPERINTENDENTS

TEACHERS:Australians. See AUSTRALIANS:TEACHERS

89 TEACHERS:Education/Sociology
Sociology as a main course. B.G. Burgess. Cambridge Inst. of Education Bul., 3 (7) 1968 p.9-10.
(Abstr. 591)

TEACHERS.,Higher education:Americans. See AMERICANS:Higher education,.Teachers

90 TEACHERS;Social integration.,Agricultural communities(Americans
Teacher participation in the community: role expectations and behaviour. R.B. Carson et al. Eugene, Oregon, Centre for the Advanced Study of Educational Administration, 1967
(Abstr. 595)

91 TEACHERS:TRAINING COLLEGES,.Students;Achievement:Tests
Predicting student performance in colleges of education. G.A. Cortis. Br. J. Educational Psychology, 38 (2) 1968 p.115-22.
(Abstr. 604)

92 TEACHERS:TRAINING COLLEGES,.STUDENTS;Teaching:Communicating.Flexibility
Predicting teaching success. A.C. Crocker. Education for Teaching, 76 1968 p.49-52.
(Abstr. 605)

TEACHERS, White;Examining:Boys(Negroes(Americans. See AMERICANS.NEGROES.BOYS:Examining; Teachers,White

93 TEACHING
Analysis of teaching behaviours. A.J. Bishop & L.B. Lavy. Education for Teaching, 76 1968 p.61-5. (Abstr. 579)

TEACHING
Related Headings:
COUNSELLING
TEACHING:16-18 Year old students.,Secondary education:Irish. See IRISH:SECONDARY EDUCATION., STUDENTS,16-18 YEAR OLD:Teaching

94 TEACHING.Aims:Opinion;Head teachers
Teacher's role: a comparative study of the conceptions of college of education students and head teachers. D.S. Finlayson & L. Cohen. Br. J. Educational Psychology, 37 (1) 1967 p.22-31. (Abstr. 620)

95 TEACHING.Aims:Opinion;Teaching students
Teacher's role: a comparative study of the conception of college of education students and head teachers. D.S. Finlayson & L. Cohen. Br. J. Educational Psychology, 37 (1) 1967 p.22-31. (Abstr. 620)

TEACHING,Bilingual:Greenlanders. See GREENLANDERS:Teaching,Bilingual
TEACHING,Bilingual:Swahili-English bilingual students.,Schools:Kenyans. See KENYANS:SCHOOLS,. STUDENTS,Bilingual,Swahili-English:Teaching,Bilingual

96 TEACHING:Equipment
Emergence of instructional systems - the educationist's predicament. J.O. Fritz. Canadian Education and Research Digest, 8 (2) 1968 p.119-32. (Abstr. 625)

97 TEACHING.Profession
Professionalism without autonomy: the paradox of modern education. C.S. Bowers. J. Educational Thought, 2 (2) 1968 p.68-77. (Abstr. 586)

TEACHING:Students;Opinion:Aims(Teaching. See TEACHING.Aims:Opinion;Teaching students
TEACHING;Students.,Training colleges:Teachers. See TEACHERS:TRAINING COLLEGES,.STUDENTS; Teaching

98 TECHNICAL COLLEGES:Leavers;Migration:Effect of educational grades
Where did the school leavers go? J.W. House, A.D. Thomas & K.G. Willis. University of Newcastle-upon-Tyne, Geography Department, 1968 (Abstr. 654)

TECHNICAL EDUCATION:Britons. See BRITONS:Technical education

99 TECHNICIANS;Migration
"Brain drains" in an integrating Europe. R. Blandy. Comparative Education Rev., 12 (2) 1968 p.180-93. (Abstr. 582)

TRAINING COLLEGES:Teachers. See TEACHERS:Training colleges
TYPOLOGY:Psychology. See PSYCHOLOGY:Typology
UNDER ACHIEVEMENT;English/Students.,Fifth forms (Grammar schools. See GRAMMAR SCHOOLS. FIFTH FORMS,.STUDENTS/English;Under achievement
UNITED KINGDOM. See GREAT BRITAIN
UNITED STATES
See
DEWEY,John.Educational theory:Adaptation;Progressive groups..United States
UNITED STATES. See also AMERICANS
UNIVERSITIES:Americans. See AMERICANS:Universities
UNIVERSITIES:Australians. See AUSTRALIANS:Universities
UNIVERSITIES:Governing bodies:Relations with students.,Universities. See UNIVERSITIES,. STUDENTS:Relations with university governing bodies

100 UNIVERSITIES/LIBERAL ARTS,.STUDENTS;Attendance:Motivation:Self assessment,Retrospective
Looking backward: how consistent are retrospective reasons for going to college. A.A. Dole. J. Counselling Psychology, 15 (3) 1968 p.263-8. (Abstr. 608)

101 UNIVERSITIES:Relations with governments
Collected seminar papers on relations between governments and universities. London. University of London Inst. of Commonwealth Studies, 1967
(Abstr. 601)

102 UNIVERSITIES:Role
Universities in the modern world. B. Fletcher. Oxford, Pergamon Press, 1968
(Abstr. 623)

103 UNIVERSITIES/Science:Research
Fundamental research and the universities: some comments on international differences. Paris, Organisation for Economic Cooperation and Development, 1968
(Abstr. 576)

UNIVERSITIES:Singhalese. See SINGHALESE:UNIVERSITIES

104 UNIVERSITIES,.STUDENTS;Maturation:Questionnaires
Institutional objectives and student development in college. A.W. Chickering. J. Applied Behavioral Science, 3 (3) 1967 p.287-304.
(Abstr. 599)

105 UNIVERSITIES,.STUDENTS:Relations with university governing bodies:Questionnaires
Students' wish for involvement. T. Blackstone. New Society, 12 (301) 1968 p.20-1.
(Abstr. 581)

106 UNIVERSITIES,.STUDENTS,WOMEN;Achievement:Correlation with child-parent relations:Picture tests
Relationship of picture structure to achievement motivation in college women. T. G. Alper & E. Greenberger. J. Personality and Social Psychology, 7 (4) 1967 p.362-71.
(Abstr. 569)

VERBAL ABILITY(CHILDREN(Slum dwellers. See SLUM DWELLERS.CHILDREN. Verbal ability

VERBAL INTERACTION PATTERNS;Governing bodies(Schools:Canadians. See CANADIANS:SCHOOLS. GOVERNING BODIES;Verbal interaction patterns

107 VOCATIONAL GUIDANCE
Life career game. S. Boocock. Personnel and Guidance J., 46 (4) 1967 p.328-34.
(Abstr. 584)

VOCATIONAL UNFITNESS(STUDENTS.,Universities:Singhalese. See SINGHALESE:UNIVERSITIES,. STUDENTS.Vocational unfitness

108 WELSH LANGUAGE:Speakers:Distribution
Language and society in Wales. R.B. Jones. Comparative Education, 4 (2) 1968 p.205-11.
(Abstr. 666)

WHITE TEACHERS;Examining:Boys(Negroes(Americans. See AMERICANS.NEGROES.BOYS:Examining; Teachers,White

WITHDRAWN STUDENTS.,JUNIOR SCHOOLS. See JUNIOR SCHOOLS,.STUDENTS,Withdrawn

WOMEN STUDENTS.,Universities. See UNIVERSITIES,.STUDENTS,WOMEN

WORD ABILITY. See VERBAL ABILITY

WORKING CLASS STUDENTS.,Public schools. See PUBLIC SCHOOLS,.STUDENTS,WORKING CLASS

YOUNG PEOPLE. See ADOLESCENTS

YOUTH. See ADOLESCENTS

COST BENEFIT ANALYSIS : PROCEDURES AND APPLICATIONS

by

M H PESTON

84/85

Introduction and methodology 87

The applications of cost-benefit analysis to the use of computer techniques in educational documentation and information 94

I. INTRODUCTION AND METHODOLOGY

1. The purpose of this section is to describe cost-benefit analysis in a general way, and to set out the broad procedure or set of procedures which must be followed independent of the particular problem under consideration. In the remainder of the paper attention will be paid to the specific cost-benefit analysis of the application of computer techniques to educational documentation and information.

2. In its barest essentials cost-benefit analysis is to be regarded as a tool of decision making. It is to be classified with (although it is not the same as) such other recent developments as operational research, programme budgeting, and econometrics. It involves both the theoretical analysis of problems, and the bringing to bear of empirical information, much of it quantitative in nature. In addition, however, it takes place within a policy or decision making environment, and is, therefore, not to be seen purely or even chiefly as a research activity. To put the point differently, cost-benefit analysis is normative in character and is inevitably involved with criteria of evaluation.

3. I shall assume in what follows that the decision making process under consideration is equivalent to one in which there is a single decision maker or a team of decision makers with the same objectives in view. What this means is that only one set of criteria of evaluation needs to be brought into the picture and that the decision making process is not self-contradictory. This does not mean that there can be no problems of discovering the value or benefit, positive or negative, of a particular activity or part of an activity, but it does mean that there is such a value to be discovered. There are organisations in which the decision makers who control them are in conflict with one another, this conflict frequently being about the objectives of the organisation. The approach to cost-benefit analyses to be outlined here ignores such conflicts, and would, therefore, not be applicable without modification to situations in which such conflicts were dominant.

4. What is a benefit or what is a cost is ultimately what the relevant decision maker agrees is a benefit or cost. This depends, therefore, on his criteria of evaluation (or utility function or preference function to use the expressions common in economics) happen to be, or on what he can be persuaded or helped to accept. It also depends on his frame of reference or universe of discourse, so to speak. The significance of an action is not independent of who takes the action or in whose interest it is being evaluated. To take the obvious example, a cost-benefit analysis of a particular action will cover more consequences of the action if it is being done for a government than if it is being done for an individual enterprise within the country.

5. The consequences of any action that falls within the frame of reference of the decision maker will be referred to as the direct costs and benefits of the action. They comprise what it is his business to take account of. The remaining consequences which fall outside his frame of reference (but inside the frame of reference of other decision makers) will be referred to as the indirect costs and benefits. Other terms for these latter are *neighbourhood effects, spillovers,* or external economies and diseconomies.

6. Frequently, the most interesting and also most difficult part of any cost-benefit analysis arises from these neighbourhood effects. Indeed, cost-benefit analysis is often motivated by governments wishing to determine the need for intervention into private sector and other decentralised decision making when these lower level decisions do not take account of spillovers. The new factory may appear a perfectly viable economic proposition if the smoke that it generates which imposes costs on other firms and households is ignored. This smoke is a spillover from the particular factory decision maker to other decision makers. The new airport seems worthwhile if the noise that it generates and adversely affects the surrounding neighbourhood is left out of account. The new underground railway may appear not worthwhile unless the consequent reduction in congestion in other parts of the transport system are taken into account.

7. Because of the existence of these neighbourhood effects it is not sufficient in judging any activity to discover whether the money revenues that may be recovered when operating it cover or more than cover the money costs of operation. The money revenues may exceed the money costs and yet the activity may not be worthwhile because of the external diseconomies which it generates. The money costs may exceed the money revenues and yet the activity may be worthwhile because of the external economies to which it gives rise. Note in both these cases the use of the word "may". An activity yielding external diseconomies may yet be worthwhile if the direct benefits exceed the direct costs by a sufficiently large amount. Equally, an activity yielding external economies may nevertheless not be worthwhile if the direct costs exceed the direct benefits by a sufficiently large amount. It is important to emphasise this to avoid any "vulgarisation" of the use of the idea of neighbourhood effects in decision making. It is not sufficient to show their existence to justify or suppress an activity. They are merely one part of the sum to be weighed in the balance with other parts.

8. Apart from neighbourhood effects there are other reasons why it is not sufficient to look at the money flow in and out of a project in order to determine its value and come to the correct decision. In a great many fields governments decide not to charge for services publicly provided. Obvious examples include education, law and order, defence, many recreational facilities and sanitation and public health. In the United Kingdom most of the National Health Service is not charged for and nor are the roads. It could be argued that some of these services should be charged for, but as a matter of principle the government decides to provide them "free". ("Free" in this case does not, of course, mean at no cost to the economy, and, indeed not at no cost to the consumer. It means merely that the consumer does not pay a money price or user charge.) This is usually the case with private and secondary education. Sometimes, however, it is impossible to charge for a service because there is no way of confining its benefits to particular users or ascertaining via the price mechanism what those individual benefits are. The typical examples here are defence and law and order. Over the relevant range the provision of defence and law and order for one or some implies the provision of these things for all. (Note that this is not entirely and necessarily the case. It is possible to provide these things privately and confine their benefits to individuals or small groups. People who buy fall-out shelters or burglar alarms are clearly doing precisely that. It is rather that most defence and internal stability are provided most cheaply on a public basis.)

9. Whatever the reason may be, the fact is that user charges do not exist for some facilities within the public sector. It is absurd to argue, however, that insofar as they involve costs, their net benefits are negative. It is still necessary to decide whether or not they should be provided, or, more generally, what is the optimum scale on which to provide them. (Incidentally, the same problem arises within firms in the private sector. Within a firm there will be activities solely devoted to assisting other activities within the firm. It is not necessarily the best way to run a firm that each part of the firm should charge each other part for the use of its services. For some firms such an internal price mechanism may be a useful form of organisation for others it may not. Where it is not, there remains for the firm the problem of deciding the optimum scale of these non-revenue producing activities.) Cost benefit analysis, therefore, has a great role to play where for one reason or another the price mechanism is not used.

10. Even where the price mechanism is used, however, decision making should not be based simply on estimates of revenue flow in and cost flow out. It has been remarked in paragraph 9 that the best way to approach the decision problem is in terms of the optimal scale of an activity. Whether or not to engage in an activity at all is then interpreted simply as whether the zero scale of operations is optimal. Consider now how costs and benefits vary as scale varies. As scale increases costs will increase and so will revenues. The increase in costs resulting from a small increase in scale is called marginal cost, the increase in benefits is called marginal benefit. (If benefits and costs are thought of as functions of scale in the mathematical sense, marginal cost is the first derivative of the cost function with respect to scale and marginal benefit is the first derivative of the benefit function with respect to scale.)

Suppose starting from zero scale we increase scale by a little amount. If marginal benefit exceeds marginal cost this increase is obviously worthwhile. Suppose we increase scale by a further small amount. Again if marginal benefit exceeds marginal cost, this increase is worthwhile. It follows that it is sensible to increase the scale of any activity if marginal benefit exceeds marginal cost. It equally follows that it is sensible to reduce the scale of any activity if marginal cost exceeds marginal benefits. Thus, the optimal scale of any activity occurs where marginal benefit equals marginal cost. (Note that this argument depends on these small variations in scale being feasible.) Now subject to a number of assumptions, it can be shown that the price which consumers are willing to pay for an article or service measures its marginal benefit. (Strictly speaking it measures its marginal private benefits, but neighbourhood effects have already been dealt with in the previous paragraphs.) Thus, the optimum pricing policy, that at which the scale of an activity will be correct, occurs at the point at which marginal cost equals prices. It can happen, however, that for certain commodities (those subject to economies of scale or economies of mass production) marginal cost is less than unit cost. It follows that a pricing policy which equates price and marginal cost while it may be optimal does not generate enough total revenue to cover total costs. Here again the cost-benefit analyst comes in to see whether a monetary deficit does arise from this cause of economies of scale or genuinely corresponds to a sub-optimal situation.

11. In fact, the position is even more complicated than that because of the so-called problem of second best. It is difficult to explain this without requiring an expert knowledge of economics on the part of the reader. Essentially it amounts to the proposition that if the remaining activities in the economy (i.e. those other than the one under consideration) are not being operated at their optimum level as specified by the equality of marginal benefit and marginal cost, the best level to be chosen for the activity under consideration is also not one at which marginal benefit equals marginal cost.

12. In a full employment situation the expansion of any activity involves the contraction of some other activity. This will represent a net gain if the benefits accruing from the former are greater than the benefits foregone by the latter. If marginal benefit equals marginal cost of the latter (i.e. the contracting activity) the switch of resources to the expanding activity indicated by marginal cost will also measure the benefit foregone by contracting the other activity. In other words, marginal cost will also indicate the marginal benefits of activities foregone. Where, however, marginal cost does not equal marginal benefit elsewhere in the economy, this will not be the case. In that circumstance the marginal cost of any activity will not indicate the marginal benefit foregone elsewhere.

13. An example concerns monopolistic elements in the rest of the economy. If a particular activity expands at the expense of another activity which is sold under conditions of monopoly, in which price exceeds marginal cost, the benefit foregone will be measured by the price. In this case the marginal cost of the expanding activity will underestimate marginal benefit foregone, and if the simple marginal cost rule is applied this activity will be operated at too high a level.

14. It may be that the activity under monopoly control is not competitive with the one under consideration but complementary to it, i.e. it may be used as an input in its production. In this case the price quoted and entering into marginal cost may exceed the benefits foregone so that marginal cost as measured is "too high". If the simple marginal cost rule is applied the activity under consideration will be operated at too low a level.

15. These second best considerations merely serve to reinforce the point that it may be, and is, indeed, likely to be the case that it will not be efficient especially from a national standpoint to determine the optimum level of operation of any activity simply on the basis of revenue and money cost considerations. If firms in the private sector do consider only revenue and costs, while maximising their profits, they may not be maximising national welfare. Similarly, while it may be appropriate to set financial objectives for firms in the public sector, these must be adjusted for all the

considerations mentioned in the preceding paragraphs. Lastly, a great deal of activity in the public sector is not carried out by firms but by government agencies for whom financial considerations are not primary. For them the need for cost-benefit analysis is even clearer.

16. Having established the basic necessity for cost benefit analysis and seen the general theoretical context in which it arises, it is now possible to examine a number of more technical matters. These are as follows:-

(a) The problem of measuring benefits.

(b) Flows of costs and benefits over time and the discount rate.

(c) Risk and uncertainty.

17. In economics the words "cost" and "benefits" are used symmetrically. A cost is merely a negative benefit. An increase in benefit is equivalent to a reduction in cost. An increase in cost is equivalent to a reduction in benefits. It follows that by implication the problem of measuring benefits encompasses that of measuring costs.

18. It has already been remarked that where a commodity is sold on the open market there is available at least a partial measure of its benefit, namely the price at which it is sold. For all the reasons already mentioned this is not a full measure, but it is a beginning. It also suggests that where a commodity or service is not sold on the open market its marginal benefit can be ascertained in the first instance by discovering what price people are willing to pay for the service at whatever rate of provision or rates of provision are under consideration. It is for this reason that much cost-benefit analysis involves either some market research to discover people's demands for various services or investigation and analysis of various proxy measures which might reveal these demands.

19. Before examining the difficulties involved in estimating benefits in this way reference may be made to the estimation of benefits as a reduction in costs. This is especially important in the present context because, as will be seen, that is the direction in which it is most likely that cost benefit analysis can assist decision making in the computer field. It may be the case that there exist a variety of alternative ways of achieving the same objective. It may also have been decided already that the objective is to be achieved. This means that the absolute benefit, so to speak, of the activities under consideration does not need to be estimated. It has been accepted *a priori* that it is above any of the possible costs. The cost-benefit analysis in this situation can now concentrate entirely on the cost side, and the decision problem becomes one of cost minimisation. It is important to bear this in mind because in many fields absolute benefit estimation appears an extremely difficult if not hopeless task. This is no reason not to examine the costs of the alternative methods of operation in the field. Furthermore, it may be possible to specify and identify the benefits of various activities without being able to measure them in the sense of expressing them in units commensurable with the costs. Nevertheless, if it is possible to recognise that the activities generate the same benefits, whether or not we have an absolute measure of those benefits, we can choose between the activities on the basis of cost alone.

20. In fact we may go further than this by recognising that many alternative activities have costs and benefits in common. It is clear that we may choose between them by eliminating or cancelling out these common elements. We may have, for example, a choice of where to locate a particular activity. In this case we are interested solely in those benefits and costs that vary with location. Similarly, if we are concerned with when to start a particular activity, our sole interest will lie with benefits and costs that vary with the starting date. A frequently occurring problem concerns the scale of operation of an activity. It is not usually the case that scale is given; rather, there exists the possibility of operating at a higher or lower level of activity. In making such a choice attention may be focussed on those parts of the activity that change as scale changes and on the way in which these variations occur.

21. Turning now to the question of market research without suggesting that this is not a useful procedure it is important to bear in mind what its limitations are. The chief of these arise from the attempt to ask questions about hypothetical circumstances. It is well known that people find it difficult to describe accurately their present situation and recent behaviour. It is much more difficult for them to give meaningful answers to questions about states of affairs which have not come into existence yet or which may be far from their normal experience. This is, of course, a matter of degree. Not all market research is about circumstances a long way from current experience, and if the research is carried out with care and subtlety, even the most unlikely situation may appear "realistic". The greatest problem is that individual preferences, attitudes, and behaviour are determined by a number of independent variables and it is extremely difficult (and expensive) to take the effects of all these into account. It is, for example, not sufficient to ask people whether or not they approve of a particular area being designated as a public park. A cost-benefit analysis requires information on how much they approve or desire the park, preferably giving rise to a measure commensurate with other costs and benefits that have to be taken into account.

22. A word must be uttered here about research in general. Partly because cost-benefit analysis is comparatively new and its practitioners are anxious (if not over-anxious) to show off its merits, partly because decision makers in the public sector are used to taking decisions hurriedly in an unprepared manner and partly because of the costs and other difficulties involved, a great deal of the work in this field has had a very thin research base indeed. Cost-benefit analysts have had to make do with published statistics which are not quite what they want. They have had to make do with informed guesses, reasonable assumptions about empirical magnitudes, and so on. A typical example is that of a local authority which wanted an estimate of people's attitudes to various aspects of a local development scheme. It obtained this by asking the members of the technical planning committee for their preferences concerning the different parts of the scheme. The present writer in connection with a computer information scheme to assist a particular profession rang up the few members of the profession he knew in order to gain an estimate of the average extent to which members of the whole profession used their books and libraries. In both these cases these extremely crude methods were employed because of the need for speed and the desire to avoid large scale expense. (In the latter case the research was only a pilot project, anyway.) The important point, however, is that no analysis can be better than its empirical underpinning, and with the best will in the world there will be costs in the sense of taking sub-optimal decisions if the appropriate research is not undertaken. Equally, of course, if the research is really too costly or involves excessive delay, it may be necessary to go ahead without it. We are dealing here with that most vexing problem, the cost-benefit analysis of research, and even with the cost-benefit analysis of cost-benefit analysis.

23. Reference has been made in paragraph 18 to the indirect estimation of benefits. In situations in which people do not reveal what value they place on a particular activity it may be possible to obtain an estimate of this valuation indirectly by examining some other part of their behaviour and making some reasonable additional assumptions. The usual example occurs in the field of transport in which a crucial benefit is the time saving to which the new development gives rise. The value of this time saving may be put at the wage rate per hour of the travellers using the new facility. More generally, any time saving development may be evaluated in terms of the income per hour of its beneficiaries. Another example occurs in the field of recreation in which the value of a public recreation facility may be estimated by the transport cost involved in getting to it. Yet again an attempt may be made to get at the value of urban amenity by looking at how house prices vary with various aspects of the environment. Similarly, the value of accident prevention in one field may be obtained be considering what people freely pay for insurance in other fields.

24. There is no shortage of examples of the use of indirect methods of estimation. It must be noted, however, that in almost all cases there are difficulties involved in using this approach. Thus, it appears that even at the margin people do not value time saved for leisure as equal to the value of income foregone, and, anyway, much of the time saved by a transport development does not occur at the work leisure margin.

Equally, in the case of recreation not all of the trip to the facility is a disbenefit. People enjoy the journey or travel long distances merely for their own sake. In the case of risk of accidents, there is some evidence that people require (albeit irrationally) all sorts of activities to be carried out more safely than they choose to behave in their personal lives. Apart from these points, there are a large number of technical problems arising as to whether a particular measure is a good approximisation to the variable it is required to evaluate. It is possible to show, for example, that people may regard environment as important and yet that this may not show up fully or at all in house price differences. It must be concluded therefore that while indirect methods of estimation of benefits may have to be used in a good many cases, they must always be used with care and in association with other methods of estimation of benefits, if possible.

25. We turn now to an entirely different aspect of the problem - the flow of costs and benefits over time. Almost all the decisions which the cost-benefit analyst is required to investigate are investment decisions in that they involve these temporal flows. Now a unit of any benefit or cost available at one point of time is not equivalent to a unit of the same benefit or cost at a different unit of time, i.e. the location in time of benefits and costs is significant. Given this it is necessary both to specify the temporal occurrence of benefits and costs and to devise a method of making them commensurable. It will be helpful in understanding this if we consider some examples. If we postpone building something (say) a school for a year, we will gain all the advantages of having some resources available for a year, but the benefits of the building will all be postponed by a year. Are these advantages which may be expressed in terms of the productive use of the available resources for a year or their use in current consumption more than offset by the postponement of the benefits of the new building? If we wish to achieve a certain end such as the production of electricity we may be able to do this in a variety of ways which vary in their capital intensity. Thus, nuclear generation of electricity requires a greater initial investment than conventional generation by coal or oil. To offset the greater capital investment there may be subsequently lower running costs. The question then becomes one of deciding how much saving in running costs in the future offsets a saving in capital costs now. A third example arises in the defence field. With a particular expenditure of resources it might be possible to provide x aircraft of a given type by 1975 or y aircraft of the same type by 1976, y being greater than x. How much smaller than y must x be before the greater number of aircraft in the later year offsets their greater value in the earlier year?

26. Apart from recognising the importance of the temporal location of cost and benefits it is necessary to answer the quantitative question of how much a unit of something is worth if it is available earlier rather than later. This question of "rates of exchange" between time periods raises a number of fundamental questions which need not be gone into here. Suffice it to say that it is accepted that a unit of resources in general available earlier is more valuable than a unit of the same resources available later. In simple terms, a £1 spent now is regarded as a larger cost now than £1 spent a year hence, and £1 received now is regarded as a larger benefit now than £1 received a year hence. More generally, any cost incurred now is regarded in general as greater than the same cost in physical units incurred in the future, and an analogous proposition can be made about benefits. Another way of putting this is that we discount the future. The reasons for discounting are, firstly, that the resources available now can be used productively to produce larger benefits in the future, and that at the margin people require some extra consumption in the future to give up the benefits of consuming in the present. (Crudely, if you can invest money to earn 5% per annum, there is no point in proceeding with a capital project that yields only 4%; and again there is no point in proceeding with that capital project if you require £105 of consumption expenditure next year to make you give up £100 of consumption expenditure now.) There are a great many difficulties involved in determining the correct discount rate. These concern how much weight should be placed on the productivity of the capital to be employed (sometimes referred to as the *opportunity cost of the capital)* and how much on the consumption side (sometimes referred to as *time preference*). It is also argued that the government in taking public sector decisions has a special responsibility to future generations and should modify the extent to which private individuals discount future benefits and cost. These matters, however, are not of primary importance within the present context. What is important is that a discount rate should be used (for example, the Treasury have

announced that the correct rate to be used in the public sector in the U.K. at the present time is 10% per annum) and that in many cases the optimum decision will vary with the precise rate chosen.

27. The fact that most decisions involve events through time leads on naturally to a consideration of risk and uncertainty. Without going into the philosophical aspects of the problem, it is apparent that the consequences of any decision are not known for certain. Because of a lack of knowledge and of insufficient research, because frequently we are plunging into new fields, because of factors not under the control of the decision maker, and for other reasons, the decision may have a variety of possible outcomes. In addition, in some cases we may have sufficient past experience to attack probabilities to these outcomes, but in other cases we may be so ignorant as to be unable to do even that. (The economist calls the former situations ones of *risk*, and the latter ones of *uncertainty*).

28. The existence of risk and uncertainty means firstly that a great many additional possibilities must be explored. The analysis of a possible activity must take account of costs and benefits in a variety of different circumstances. In addition, some allowance may be made for the difference in risk between different activities. Thus, if an activity may yield alternative benefits at a particular point of time with various probabilities, it may not be correct merely to work out the average of these benefits and use that figure in the analysis. Instead, because of aversion to risk on the part of the decision maker some further deduction from the benefits (or addition to the costs) may be made as a risk premium.

29. It is sometimes suggested that risk may be expressed as a single figure such as the variance of the outcome of a particular decision. This can then be compared with the figure of mean benefit in order to arrive at the correct decision. The difficulty with this is that the variance may not be calculable in practice, and even when it is it may conceal important information. As a practical procedure, therefore, it may be more useful to specify the possible outcomes together with a broad idea of the probability of their occurrence, and leaving the decision maker himself to make due allowance for risk.

30. To summarise this section of the paper a cost-benefit analysis involves the following steps:-

(a) The specification of the alternative actions open to the decision maker.

(b) The prediction of the outcomes of those actions.

(c) The evaluation of the positive outcomes as benefits and of the negative outcomes as costs.

(d) The conversion of as many of these values as possible into a common unit of evaluation, frequently a unit expressed in money terms.

(e) Where the outcomes of a decision occur through time the conversion of these evaluations as far as possible into a present value using a discount rate.

(f) Where the outcomes involve risks and uncertainty, the specification of the various possibilities and an estimation of their relative probabilities.

The final picture presented will then show to the decision maker the choices open to him, an evaluation of their consequences, and a specification of the risks and uncertainties involved. It is on this basis that he must take what he considers to be the optimal decision.

II. THE APPLICATIONS OF COST-BENEFIT ANALYSIS TO THE USE OF COMPUTER TECHNIQUES IN EDUCATIONAL DOCUMENTATION AND INFORMATION

31. The purpose of the second part of this paper is to indicate how these techniques may be applied to a particular problem. Clearly, it will only be possible to indicate the applicability of cost-benefit analysis in theory, and what follows is no substitute for carrying out an actual application in this field. Moreover, it is highly likely that an actual empirical study would bring to light all sorts of special problems which will not have been thought of *a priori*. It is equally necessary to mention that some of the theoretical problems mentioned in the previous paragraphs may not be relevant to the particular problem under consideration.

32. Let us start by considering what the alternatives are. Unfortunately, no complete specification of these exists. A partial listing would include the following:-

(a) The status quo or initial situation in which computers are not used for documentation and information,

(b) A new situation in which a system of documentation and information is partially computerised,

(c) A new situation in which a system of documentation and information is fully computerised.

33. In essence there is almost an unlimited number of variations of what sort of information relevant to a particular subject can be stored in the computer, and what is the nature of the access to the computer. In its most limited form the computer may be simply the equivalent of a library catalogue with access according to author and subject matter. It may at the same time give rise to or print out particular subject catalogues. In its most extensive form the computer may repeat the most advanced form of abstracting service able to print out a summary with complete references and cross-references of all available material on any topic. Again the computer may provide a regular abstracting service, and it may, in addition, provide special abstracts on demand.

34. Along another dimension the delay time between the publication of a document and its inclusion in the computer system may be variable. Equally, the computer may cease to print out obsolete information after a particular amount of time has elapsed. An important trade-off given the aggregate size and expense of the installation may be between the degree of detail of the information provided and its "up to dateness".

35. A variation of the utmost significance concerns the size of the system as such. The larger the system the more information it can store and produce, this increase being partly in terms of coverage within a field, partly in terms of the degree of detail, and partly in terms of the number of fields that can be covered simultaneously. The importance of the last of these possibilities is strongly connected with all the inter-disciplinary developments that are currently taking place in both the natural and the social sciences. Equally, since it may be expected that the potential number of cross-references increases multiplicatively as the number of subjects covered is expanded, the attempt to deal with inter-disciplinary matters will throw a specially heavy burden on the storage capacity of the system.

36. Another variation relates to the degree of technological advance and potential obsolescence of the system introduced. Without examining the whole of the economics of research and development the following considerations are worth bearing in mind:-

(a) Postponement of a project leads to the possibility of installing more advanced equipment subsequently,

(b) Bringing forward the project assists the learning process and may itself give rise to technical advance,

(c) Investment in less durable equipment may shorten the time interval between reinvestment in more advanced equipment,

(d) Investment in more durable equipment may lead to maximum exploitation of the existing technology.

37. It is important to bear in mind that the value of computerisation will not be independent of the subject under consideration. In other words, the optimum policy in one field (say) physics will not necessarily be the same as the optimum policy in another field (say) sociology. (Of course, what is learned about computerisation in one subject will spin off and be of value to others.) One may expect that the costs of putting the information into the computer will vary between subjects, and so will the nature of the access required and the benefits generated. In particular, in the social sciences, if one may use such a crude dichotomy as that between theoretical and factual material, one would expect the optimum decision to depend to a considerable degree on the theory-fact mix.

38. This leads on to the question of the storage of statistical and similar material. There is a significant difference between the storage of the contents of the published work of a particular subject, and the storage of its subject matter as such. Although it may be agreed that this is partly a matter of degree, a storage system which contains all the major quantitative information, time series and cross-section, would be of considerably greater value than one which merely contained information on what work had already been carried out on that information. In addition, insofar as most econometric and similar qualitative work in other social sciences is being carried out using computers, the previous computerisation of the data centrally will be of added value.

39. Let us assume now that the possible choices which have to be compared have been specified. The next stage is to make sure that the comparisons are of like with like. This is espeicallly important with regard to the size of the system and its coverage. It is not sensible, for example, merely to compare an extremely large scale fully computerised system with a small scale conventional system. It is equally not sensible simply to compare a computerised system without individual access with a conventional system with individual access.

40. A much more useful procedure especially in the initial stages of exploration of this whole field is to take a particular system, work out its full costs and benefits and then to ask two questions:-

(a) Given the costs of this sytem is it possible to use them in some other way so as to generate more benefits than this system does?

(b) Given the benefits of this system is it possible to determine an alternative system which generates them at less cost?

In other words, we may standardise on the input side and maximise output, or standardise on the output side and minimise cost.

41. An important reason for this kind of standardisation or normalisation is that it helps to ensure that the relevant alternative choices are brought into consideration. It may well be that a large computerised system yields larger net benefits (i.e. the difference between gross benefits and costs) than does a small non-computerised system. It could be the case, however, that both a smaller computerised system and a larger non-computerised system could yield larger net benefits still.

42. It has been pointed out in the introduction that the typical problem we have to analyse is one of capital investment. The decision we take now involves a commitment over time and will constrain any future decisions we take. An important technical difficulty that arises in this connection concerns the length of the time horizon. If we are investing in a complex of capital equipment which is expected to last *n* years, it is not usually sensible to compare this with an alternative complex which will last a different number of years. Although this depends on the particular problem under consideration, the most useful procedure is to standardise with respect to the time horizon

as well as in the ways indicated in paragraph 40. Otherwise the comparison may be biased in favour of the asset with the longer life, whereas it is important to ask with respect to the asset with the shorter life, what happens when that life is over. (It is worth mentioning here that there is a further problem of investment decision making that is relevant to all this, namely, the question of the optimum length of life of an asset and whether to replace it. We can assume that this problem is encompassed by our analysis if one of the available alternatives is the continuation of whatever system already exists. Incidentally, in comparing an existing arrangement with something new it is important to consider only the costs that will emerge in the future and not the expenditure paid out in the past. Thus, the past training costs of personnel currently employed in a conventional educational documentation system are not to be included in the costs of this system to be compared with a computerised system that might replace it. What should be counted are, of course, any future training costs that have to be incurred.)

43. Having specified the alternatives, the next stage is to specify the constraints on the decision maker. The crucial constraint concerns the availability of capital. If the decision maker is limited by the total quantity of capital he can invest in any time period, (i.e. he is subject to capital rationing) this will immediately exclude certain extremely large projects or combination of projects. In particular, there may be available alternatives with positive net benefits and not incompatible with any other projects that still have to be excluded from the capital investment plan.

44. Of course, the capital constraint may not be one of quantity, but merely one of price. The decision maker may not be free to choose his own discount rate but will have this imposed on him by others in the same organisation or in some other relevant organisation. Thus, in the public sector in the U.K. the Treasury say that discounted cash flow methods should be used for investment decision making, and themselves lay down what is called the test rate of discount. This is supposed to measure the return that can be earned on the capital in the rest of the economy, notably the private sector, and has recently been raised from 8 *per cent per annum* to 10 *per cent per annum*. (What this means is that £100 invested this year must yield benefits worth at least £110 next year or £121 in two years' time etc. or any equivalent combination over a number of years.)

45. In discussing the discount rate one further complication may be referred to. A number of the projects under consideration in this field involve international co-operation. It also appears to be the case, however, that different countries value capital at the margin at different rates. It is by no means obvious what rate should then be used in a cost-benefit analysis. Suffice it to say that it would be necessary to agree on a rate, and the higher the rate the less capital intensive will be the optimum system.

46. In any field in which great technical progress is occurring the relevant constraint may not be one of capital but of labour especially of particular kinds of high level manpower. Once again the size of the system introduced or the number of possible projects that can go ahead at any point of time may be limited by this shortage. (It is human capital that is rationed rather than capital in the ordinary sense of the word.) Just as in paragraph 43, therefore, in this case too there can occur projects with positive net benefits that have to be foregone.

47. A decision not to engage in a project this year is not the same as a decision not to engage in it ever. Postponement of any project must always be included in the list of relevant alternatives. A project may be postponed either because of the resource limitations already mentioned or because if it is postponed there will be either a significant increase in its benefits or a significant decrease in its costs.

48. If in a situation of resource limitation a project has to be postponed that project should be deferred which suffers least or gains most from such a decision. Suppose a decision maker is confronted with n projects $P_1 \ldots P_n$. Assume that they are ranked in order so that if they are carried out this year P_1 is the best and P_n is the worst. If one has to be postponed, does this mean that this should be P_n? The answer to this question is obviously, no. Although P_n is the least favoured project this year,

it may suffer most from postponement. It could well be the case that P_1, this year's best project, would suffer least by postponement and should be treated accordingly.

49. The next stage in the cost-benefit analysis will be to predict the costs and benefits of each alternative, each cost and benefit being related to the time period in which it occurs. In the case of a computerised system the following costs are relevant:-

(a) The initial capital cost of the installation. This will include the cost of the equipment together with the installation costs of getting it into working order.

(b) The initial cost of storing the first set of information so as to make the system at all useful.

(c) The initial cost of getting the information provision system into working order.

(d) The initial cost of training personnel to get the system into working order.

(e) The running cost of the computer over time.

(f) The costs of adding the new information to the system over time.

(g) The costs of generating and printing bibliographies, abstracts etc.

(h) The costs to the system of providing special abstracting services.

(i) The cost, either as an initial capital cost or as a running cost over time, of the rent of the building in which the system is housed.

50. In the case of a non-computerised system a similar list of costs is relevant. These will include:-

(a) The initial capital cost of the installation. There will obviously be equipment involved even in a conventional system. There will be the capital cost requirement for storage space, for catalogues etc.

(b) The initial cost of storing the documents, and of cataloguing them.

(c) It is probable that there will be no special costs of getting a conventional system into working order or of training personnel to operate it, but if there were, these too will have to be taken into account.

(d) The costs, chiefly of staff, of running the conventional system.

(e) The costs of adding to the conventional system.

(f) The costs of generating and printing bibliographies, abstracts etc.

(g) The costs of providing special services.

(h) The cost of the building in which the installation is housed.

51. These cost headings have been arranged so that both systems have items under most headings. Although it is not tremendously fruitful to speculate over the relative magnitude of these items, it is worth making a few remarks about one or two of them.

52. It may be expected that the capital cost of the computer installation is higher than that of a conventional installation. Certainly, this will be true of the operating equipment required. It may also be expected that the cost of the building will be higher for a conventional system since it will have a larger physical volume of

material to store. Concerning this material, if it has to be purchased by the conventional system and only borrowed by the computer system, it too will weigh more heavily in the former's accounts than the latter's. It has already been remarked that the "settling down" costs must be expected to be higher in the case of a more technologically advanced system. The running costs of the computer system should be significantly lower than that of the conventional system. Basically once the information is in the computer it can be programmed so that no difficulties or delays are involved in gaining access to it (at least when it is operating at less than full capacity). Its ability to print out on demand will also be vastly superior to that of a conventional system. A conventional system may include a highly complicated cataloguing system of references and cross-references, but it will still be a large operation to obtain access to bibliographic and other material on any subject.

53. To summarise, it appears likely that on the cost side the essence of the decision will be a trade off between capital cost and running cost. To provide the same service it may be expected that the initial capital of a computerised system will be larger than that of a conventional system but its running cost will be lower. The problem is then to decide whether the saving in running cost is worth the extra capital cost. This saving may be expressed as a rate of return on the extra capital involved and compared with whatever rate of return is required to be earned in this sector of the economy. Alternatively, the discounted present value of the savings in running cost may be calculated at the relevant discount rate and compared with the extra capital cost.

54. Whichever approach is used there are a number of additional factors to be borne in mind. Firstly, there are risks and uncertainties to be taken into account. The capital cost of the computerised system, especially if it is an extremely advanced one, may be difficult to estimate. Equally there may be difficulties in estimating the time it takes the system to settle down and work satisfactorily. Equally, technical advance may take place within the conventional system during the life of the computer system. The productivity of the manpower employed will rise if, for example, new techniques of cataloguing and of obtaining access to material may appear. For all these reasons it may be thought desirable to impose a more stringent requirement on the computer system. If it is true that it is riskier than the conventional system some allowance must be made for this risk, by, for example, adding something to its capital cost as a risk premium.

55. A second consideration concerns the behaviour of labour costs over time. It may be expected that in advanced industrial economies the price of labour will rise relative to that of capital over time. In estimating the running costs of a system, therefore, some allowance must be made for this. In the case of the conventional system, for example, the running costs should not be estimated simply at this year's wages and salaries, but every year these should be calculated at the expected real wages and salaries pertaining to them.

56. An analysis of the costs of alternative systems is an obvious first step in any work that has to be done. If possible this should not be carried out for one size of system, but for a variety of different sizes. The reason for this is that it cannot be assumed, and is unlikely to be the case, that the costs of a system vary proportionately with its size. Thus, while it might be true that a computerised system of a given type costs less than a conventional system if both are extremely large, the reverse may be true of a smaller scale of operation. Since one of the problems will be to decide the optimum scale of operation, there is the obvious necessity of estimating costs at different scales. It follows that under the headings of paragraphs 49 and 50 the question may be asked in each case, how does this cost vary as the size of the system varies? (Since systems vary not only quantitatively but also in the nature of the service they provide, an additional question would be, how does cost vary as the service provided by the system varies in this or that way?)

57. The costs of any system are not simply those involved in its operation as a producing unit. There are also the costs of using the system as a consumer of its services. One of the benefits of any system compared with another would be the saving in time and trouble that it accords to those who use it. The consideration of these users' or consumers' costs, therefore, provides us with a direct path to the analysis of the benefits of the system under discussion.

58. One possibility is to standardise for benefits and assume that both systems involve the same amount of work and time for the user. Thus, if a computerised system is able to offer a bibliography of a given type on demand with a time delay of so many minutes, it may be compared with a conventional system offering exactly the same service. While not rejecting this altogether as a suitable question to ask, it differs somewhat from the practical question that is usually asked. This assumes that the service provided by the computerised system is quicker than that offered by the conventional system and involves a lower cost for its user. In this case the relevant comparison would be between the extra cost of the computerised system and the value of time and effort saved by its users. To estimate this latter figure experiments must be carried out to compare the time required to obtain various bibliographies, abstracts, set of references etc. It is then necessary to discover who the people are who are involved as consumers doing this. Their time may then be valued at, for example, their hourly wage rate. The resulting sum will then be the time saved multiplied by its value multiplied by the number of times this occurs per annum.

59. The relevance of time saved is not merely to be seen in terms of having the time available for something else. To be able to do something more quickly has a value over and above that, and is concerned with timing rather than time. The benefits of timing are obvious in any competitive situation in which the winner is the one who finishes first. They are equally obvious in industrial and commercial situations where ability to finish first or act earlier than anyone else may make all the difference between profit and loss. In the case of bidding for a contract the ability to get together a large amount of information rapidly may be crucial in determining the accuracy of the bid and the chance of gaining the contract profitably. In the legal profession winning or losing a particular case may depend on rapidity of access to a complete set of relevant references. What is required, therefore, is a further study of users in terms of the value to them of speed of access to information. This is an obvious example of where market research comes in, its objective being to discover who the users of a computerised system would be and how far they value speed of information provision.

60. So far we have dealt with two kinds of benefits of a computerised system. These are:-

(a) The likelihood of a considerable saving in costs compared with a conventional approach if the system is large enough, and

(b) The benefits of a saving in user time and of increased speed of access to information.

It is probable that these will account for most of the benefits to be compared with the initial investment. In other words, it is probable that a cost-benefit analysis of a computerised system will favour such a system if the need is for something on an extremely large scale, if users' time saving is of great significance, and if delays in information flow are costly.

61. There are, however, certain other benefits to be taken into account. Firstly, as has been remarked earlier there will be a spillover from the experience of an initial project to all subsequent projects. The initial investment and settling down costs of this project will provide information relevant to later projects, the costs of which will be lower than they otherwise would be. This cost saving is, obviously, a benefit attributable to the project under consideration. Another way of looking at the matter is to say that some of the costs of this project are really attributable to later projects.

62. Secondly, it may be that it will be possible by means of computer links to arrange direct access from user centres to the central computer installation. This greater ease of access, although it could be interpreted as coming under the general heading of time saving, would be a considerable advantage accruing to a computer system. Apart, however, from this advantage, there will be a further advantage in terms of an increase in the average rate of use of the system relative to its capacity. Essentially any system of this kind represents a capacity or potential available for use. The easier

the access the greater the rate of use relative to its potential. It follows that with two systems of equal capacity, the one with greater access will yield a larger flow of benefits than the other.

63. The first part of any cost-benefit analysis must be concerned with the benefits of the output of the system as such. So far we have taken these for granted, presumably on the grounds that educational documentation services already exist, and that this is prima facie evidence to support the view that their benefits outweigh their costs. Nonetheless, it cannot be taken for granted that bibliographies, abstracts etc. have a value in excess of the costs of producing them, and it will be necessary to carry out investigations to approximate this value.

64. The most obvious method is, of course, to price them at their marginal cost of production, and to determine what the demand is for them at that price. Before any computerised system comes into operation it will be possible to determine from existing systems what people are prepared to pay for various services and products. This, however, will not be sufficient partly because many of these things are currently unpriced, and partly because the new system will produce new outputs of greater scale and complexity and higher quality than anything existing today. Here again, therefore, a need for market research emerges. It will be necessary to ask potential users what value they place on these services, and especially what they would be willing to pay for them. This is not an easy matter to do, and the answers may be quite misleading if, in fact, people do not expect to pay for the service. In those circumstances they will be most tempted to exaggerate the value of the system.

65. In sum, where any item of service of a computerised system is currently being produced by conventional methods the cost to the consumer of the service is in the first instance a measure of its benefit. This is true whether the cost is expresssed as a price or as the time and trouble incurred by the consumer in obtaining the service. If, for example, people are currently compiling bibliographies for themselves the amount of their own personal input involved is certainly a lower limit to the value of the bibliography. This personal input can be measured in terms of hours spent multiplied by the value of those hours. (It is, incidentally, a lower limit because there may be other beneficiaries than the compiler, if for no other reason.)

66. Thus, the major area of uncertainty remains the value of the new services that a computerised system provides. As has been remarked in paragraph 64 market research may help to provide an answer here. But it is apparent that for those services which are sufficiently different from what is presently available market research has its drawbacks, and yet no other obvious approach is available. It is here, therefore, that judgement is most likely to have to be exercised.

67. The conclusion of this paper, therefore, is that it is quite feasible to apply cost-benefit methods to this problem. Indeed, there is very good reason why care should be taken to estimate both the capital and running costs of alternative projects. In cases where the benefits provided by the projects are the same this will provide sufficient information to take the optimum decision. It may also be possible to evaluate benefits especially when these are expressible in terms of the time saved by the users of the service. The major difficulties arise in connection with other benefits especially when these are very different from those made available by existing systems. Systems will differ in the nature of these new benefits and in the extent to which they make them available. While a value may be attached to these benefits by surveys applied to their potential receivers, such techniques are likely to be imperfect at best. It follows that this is the area in which judgement is most needed. The judgement will, of course, be exercised on the basis of all the information made available in the cost-benefit analysis. Above all the costs of these new benefits must be laid out clearly so that those who have to take the decision can decide how much extra cost is worth how much additional benefit.

COST ESTIMATES FOR BIBLIOGRAPHICAL SEARCHING IN A SOCIAL SCIENCE INFORMATION SYSTEM

by

G K THOMPSON

Introduction 103

ILO system (technical considerations) 103

ILO system (economic considerations) 107

Information networks 110

This paper was originally prepared for presentation at the Second International Conference on Mechanised Information Storage and Retrieval Systems, Cranfield, 2nd - 5th September, 1969, and is reproduced here by kind permission.

I. INTRODUCTION

1. Many of the large-scale mechanised information systems in operation today grew up because of the need of the military and para-military establishments to have rapid access to large quantities of scientific and technological information. This paper will deal with two complementary efforts in the field of social sciences, where an awareness of the information problem is acute, and the global needs staggering. The target audiences for these two systems are on the one hand social science researchers and on the other policy planners and makers in the broad field of economic and social development.

2. An operational system (in the International Labour Office) will be described, and an array of figures representing actual or theoretical operating costs will be given. As could be expected, however, these figures relate only to costs of systems development, input preparation, file maintenance and retrieval. No attempt can be made to evaluate in other than general terms the economic impact that the existence of such a system may have. It can be readily demonstrated, however, that computer-assisted information systems may make sense within individual institutional environments, and make greater sense when on-line systems linking various collaborating institutions can result in enabling each to operate more efficiently and at lower cost. The existence of such on-line networks would make the greatest sense for users because of the possibility of having access at a reasonable price to large information stores.

II. ILO SYSTEM (TECHNICAL CONSIDERATIONS)

3. The ISIS system (Integrated Scientific Information Service) was created as a nucleus of a general information system. To date the system has been used to record and retrieve information about documents and about research projects. At present (Sept. 1969) some 31,500 bibliographical records exist on magnetic disc and are being used to generate a variety of printed indexes (1) and for document retrieval. Retrieval is done either in batch-processing mode or by direct interrogation via an IBM 1050 terminal. Figure 1 shows the various stages of development of the retrieval system within the International Labour Office.

Figure 1 - Evolution of bibliographical retrieval system

	Number of records available	*Average search time*	*Average users' waiting time*
Phase 0 (1964) Planning and experimentation	---	---	---
Phase 1 (Jan. 1965-April 1969) Punched cards, incl. inverted descriptor file	0-28,000	up to 8 hours	2-3 days
Phase 2 (May 1969-) IBM 360/30 batch processing	28,001-	90 seconds	less than 24 hours
Phase 3 (Aug. 1969-) IBM 1050 terminal on-line throughout the day	31,000-	instantaneous	few minutes to few hours (depending on print-out)
Phase 4 (1970-) Visual terminals in various locations	40,000-	"	"
Phase 5 (1971/72?-) Inter-agency network	100,000-	"	"

---

(1) Examples of two of these are available in: *Tabledex and 100 Index*, Geneva, ILO, 1968, 5 p. (LD/NOTES/33)

4. Bibliographic records consist of a physical description of the document (date and language of publication, title, author, report number and so on) together with a subject abstract written in natural language. The abstract contains in the running text one or more tagged descriptors chosen from an indexing vocabulary of approximately 1200 terms. The documents chosen for recording in the system include new books received in the library, a selection of important journal articles, internal documents, technical reports or any other type of document deemed important enough to record.

5. Each record thus selected and analysed is then typed on a Flexowriter and proofread. When the paper tape is re-run for correction, the Flexowriter is coupled to an IBM card punch so that punched cards are obtained. The corrected typewritten Flexowriter manuscript is then sent for offset reproduction in two versions. Reduced photographically by 30 per cent, a weekly current-awareness bulletin containing four records per page (see figure 2) is produced together with a bristol edition which is cut up to provide standard 3"x5" (75x125 mm) catalogue cards. This weekly bulletin, *International Labour Documentation* is distributed to approximately 1,000 recipients, ILO research staff or outside libraries or institutes. The catalogue cards are filed in two locations in Geneva, forwarded to the Union Catalogue housed in the National Library in Berne and also supplied on request to ILO researchers wishing to keep their own manual files.

6. Once a week the input data on cards are corrected by a number of check programs and reformatted into a structure similar to MARC II. The computer produces an alphabetical subject index which is published in each weekly issue of *International Labour Documentation* and updates the master file and inverted descriptor file. The complete master record is stored in duplicate on magnetic tape and a working record on 2311 discs (2314 in April 1970). Various cumulative indexes can be prepared from this record as required.

7. On-line retrieval programs, written in Assembler, are intended for use in the foreground partition of an IBM 360/30, and require 18K core storage. The normal procedure is to first formulate the search questions directly on an IBM 1050 typewriter terminal. The simplest strategies permit searches by descriptors, language and date of publication in Boolean combinations of *ands*, *ors* or *nots* (see figure 3). As each new element is typed on the terminal keyboard, the computer responds by indicating the number of records containing that element, together with the number of matches with previous elements in the search. At any point in the formulation the computer may be asked to display a few matches to see if the search is proceeding along desired lines. If the user can be satisfied with abstract numbers only or if the number of hits is very small, the search may be dealt with entirely on the terminal. If, however, he requires the complete bibliographical records, a note is made by the terminal retrieval programme and printing is done on the high-speed printer in batch mode at designated hours during the day.

8. During the first year of operation (1965) the system ran in parallel with the old library system, each item being catalogued and classified twice. When it was decided that the new system could supplant the old one, it was felt nevertheless that the manual card catalogue should be maintained. However, the old dictionary catalogue was abandoned in favour of separate author, title, subject, geographical and report number files (2).

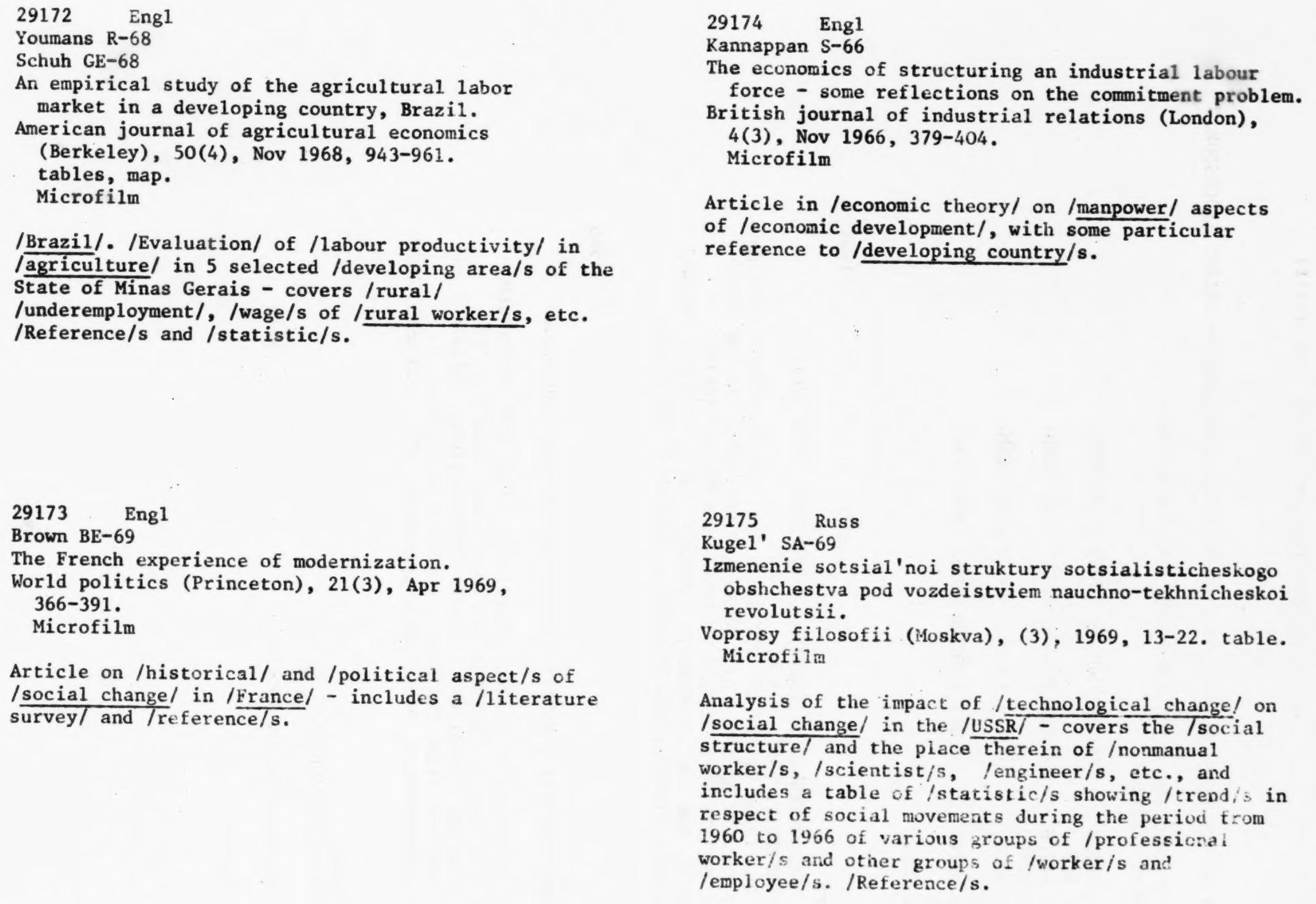
9. In the early days it was found useful to take advantage of existing IBM KWIC programs rather than to create an input format requiring new software. The KWIC format also proved useful for an unexpected reason: the over-all record length and number of permissible characters per line corresponded exactly to what would fit on a

---

(2) Laraine Kenney: "The Implications of the Needs of Users for the Design of a Catalogue: A Survey at the International Labour Office", *Journal of Documentation*, vol. 22, no. 3, Sept. 1966, pp. 195-202.

Figure 2

Sample page from weekly bulletin "International Labour Documentation"



29172 Engl
Youmans R-68
Schuh GE-68
An empirical study of the agricultural labor
  market in a developing country, Brazil.
American journal of agricultural economics
  (Berkeley), 50(4), Nov 1968, 943-961.
  tables, map.
  Microfilm

/Brazil/. /Evaluation/ of /labour productivity/ in /agriculture/ in 5 selected /developing area/s of the State of Minas Gerais - covers /rural/ /underemployment/, /wage/s of /rural worker/s, etc. /Reference/s and /statistic/s.

29173 Engl
Brown BE-69
The French experience of modernization.
World politics (Princeton), 21(3), Apr 1969,
  366-391.
  Microfilm

Article on /historical/ and /political aspect/s of /social change/ in /France/ - includes a /literature survey/ and /reference/s.

29174 Engl
Kannappan S-66
The economics of structuring an industrial labour
  force - some reflections on the commitment problem.
British journal of industrial relations (London),
  4(3), Nov 1966, 379-404.
  Microfilm

Article in /economic theory/ on /manpower/ aspects of /economic development/, with some particular reference to /developing country/s.

29175 Russ
Kugel' SA-69
Izmenenie sotsial'noi struktury sotsialisticheskogo
  obshchestva pod vozdeistviem nauchno-tekhnicheskoi
  revolutsii.
Voprosy filosofii (Moskva), (3), 1969, 13-22. table.
  Microfilm

Analysis of the impact of /technological change/ on /social change/ in the /USSR/ - covers the /social structure/ and the place therein of /nonmanual worker/s, /scientist/s, /engineer/s, etc., and includes a table of /statistic/s showing /trend/s in respect of social movements during the period from 1960 to 1966 of various groups of /professional worker/s and other groups of /worker/s and /employee/s. /Reference/s.

Figure 3

Sample terminal search

(Question: Documents on the ILO written in Russian, but excluding publications of the ILO itself)

```
  BONJOUR -- THE TERMINAL MONITOR SYSTEM IS NOW AVAILABLE -- SELECT PROCESSOR.
SEARCH
YOU HAVE REACHED THE QUICK AND DIRTY SEARCH ROUTINES.
ENTER SEARCH QUESTION.
= ILO
00138 NOW MATCH      00138 INDEXED UNDER LAST TERM.
* ROLE OF ILO
00451 NOW MATCH      00332 INDEXED UNDER LAST TERM.
* ILO MENTIONED
01089 NOW MATCH      00639 INDEXED UNDER LAST TERM.
□ ILO PUB
00916 NOW MATCH      01063 INDEXED UNDER LAST TERM.
LANG=RUSS
00029 NOW MATCH
PRINT 2

   25222                                                     1967
PIMENOV PT
ISTORIIA PROFSOIUZNOGO DVIZHENIIA ZA RUBEZHOM, 1939-1966.
MOSKVA, PROFIZDAT, 1967. 526 P.
   STUDY OF /TRADE UNION/ ACTIVITIES IN THE /INTERNATIONAL/ SPHERE
   AND IN /EUROPE/AN AND /LATIN AMERICA/N COUNTRIES AND THE /USA/
   DURING THE PERIOD FROM 1931 TO 1966 - COVERS /SOCIALIST
   COUNTRY/S, THE /UK/, /FRANCE/, /GERMANY (FR)/, /ITALY/, /CUBA/,
   THE /WFTU/, /IFCTU/, AND INCLUDES A CHAPTER ON THE /ILO/.
RUSS

   24953                                                     1968
VASILEV AA
O SOVERSHENSTVOVANII I RAZVITII POLITEKHNICHESKOGO TRUDOVOGO
OBUCHENIIA.
SHKOLA I PROIZVODSTVO (MOSKVA), 12(7), JUL 1968, 3-8. MICROFILM
   /USSR/. ARTICLE ON PROBLEMS OF /VOCATIONAL TRAINING/ AND
   POLY/TECHNICAL EDUCATION/ - COVERS /PREVOCATIONAL TRAINING/ AND
   /PART TIME//IN PLANT TRAINING/ OF PUPILS OF /GENERAL
   EDUCATION//SCHOOL/S, /VOCATIONAL GUIDANCE/, ETC. /ILO MENTIONED/.
RUSS
$
END OF SEARCH.
ENTER SEARCH QUESTION.
EXIT
TERMINAL CLEAR -- SELECT PROCESSOR.
STOP
TERMINAL CLOSED.
-
```

standard library catalogue card using the technique described above. As time went by, it was found that the KWIC format did not offer sufficient versatility either in tagging the various data elements or in file manipulation and storage within the computer. A number of minor modifications were subsequently made, for example to permit tagging of report numbers and language designators.

10. Retrieval on collators and sorters began in October 1965 after approximately 5000 records had been accumulated. During the period October 1965-August 1969, some 1200 bibliographical searches were carried out, the number of requests going up sharply, as could be expected, after the switch-over from punched cards to the IBM 360/30 in May 1969. Experience with the first 1,000-odd searches revealed that certain groupings of terms, either geographic or subject, frequently emerged from the questions that had been asked. This has led to the creation of 40 "ANY" tables. These ANY tables are in an informal sense clusters or hierarchical groupings, but they are not part of the indexing vocabulary and are only used in retrieval (see figure 4).

Figure 4 - ANY tables

Geographic

Any Africa
Any Africa, South of Sahara
Any Arab country
Any Asia
Any Caribbean
Any Central America
Any Comecon country
Any Developed country
Any Developing country
Any East Africa
Any EC country
Any Europe
Any Far East
Any French-speaking Africa
Any Latin America
Any Mediterranean country
Any Near and Middle East
Any North Africa
Any North America
Any Oceania
Any OECD country
Any Role of developed country
Any Scandinavian country
Any Socialist country, Asia
Any Socialist country, Europe
Any Socialist country, World
Any South America
Any South East Asia
Any Southern Africa
Any West Africa

Subject

Any Agriculture
Any Education
Any ILO
Any Labour-management relation
Any Labour market
Any Research
Any Rural
Any Service Ind.
Any Vocational training
Any Wage

11. It is estimated that some 50,000 bibliographical records of 500 characters average length can be stored on one 2314 disk, together with the inverted descriptor file and the ANY tables. An analysis of the frequency of occurrence of all elements in 15,000 records containing more than three characters revealed that data compression techniques judiciously applied could reduce the length of the master file by 30 per cent. Less frequently consulted records, or records segregated by some other criteria, such as date, language or subject, could be stored off-line on additional disks. These disks could be changed manually on request, and in this way a number of records well in excess of 50,000-65,000 would be accessible within minutes.

III. ILO SYSTEM (ECONOMIC CONSIDERATIONS)

12. Developmental costs for the system include nine man-years systems design, analysis and programming, at an average cost over the past five years of $16,000 per man-year, or a total of $144,000. To this must be added $25,000 for the rental of outside computers for testing purposes. Overhead expenses have not been included, but would comprise general administrative management staff costs, rental of premises, heat, lighting, purchase of equipment (other than EDP), office supplies, etc.

13. The unit cost for preparing a bibliographical record in machine-readable form is $6.50 calculated at the 1970 standard cost factor in use in the ILO. This figure includes staff time for descriptive cataloguing, document analysis, editing, flexowriting and proof-reading and is based on current productivity figures taking into account an average of 18.5 working days per month. Bearing in mind the increases in the salary scale over the past five years, an average unit cost for the entire period is about $5.75. Therefore, the total data preparation costs for the 31,500 records were $181,125, excluding overhead, or an annual expenditure of $38,820.

14. Annual hardware rental costs for peripherals needed to support the bibliographical work can be estimated at $10,000. The computer would be used approximately one hour per week for file maintenance, check programs and updating. Although the computer centre does not charge the Central Library and Documentation Branch for use of the facilities, hardware and computer time for updating and file maintenance have been reckoned at some $15,000.

15. Before attempting to calculate a hypothetical cost for performing retrieval on the computer, it is first necessary to allocate over-all developmental and other costs to the various types of product at present being obtained from the system. A possible breakdown might be:

Weekly current awareness bulletin - 20%

card catalogues - 20%

book catalogues and other printed indexes - 20%

retrieval - 40%

100%

To arrive at annual costs, it is further proposed to distribute developmental costs over a ten-year period. The over-all data preparation costs must also be taken into account, and since it can be assumed that one would want to search the whole file over a ten-year period, this cost is also split up in a similar manner. All other factors being equal, it can be assumed that the cost per search would depend on two variables - the size of the file and the number of searches performed each day (see figure 5).

Figure 5

Cost of performing searches on the computer

	per year	*per search* (15/day)	*per search* (20/day)	*per search* (40/day)
a) Developmental cost (1/10th of 40% of $169,000)	$ 6,760	$ 2.03	$ 1.52	$ .76
b) Data preparation cost (1/10th of 40% of $181,125)	$ 7,245	$ 2.18	$ 1.63	$ .82
c) Hardware cost (40% of $15,000)	$ 6,000	$ 1.80	$ 1.35	$ .68
d) Computer time (CPU) (12 minutes/search)		$ .60	$ .60	$ .60
e) Staff time to formulate questions on terminal		$ 1.40	$ 1.40	$ 1.40
f) Computer time to print results		$ 1.00	$ 1.00	$ 1.00
		$ 9.01	$ 7.50	$ 5.26

16. Although some of these figures should probably be taken with a grain of salt, they are indicative of an order of magnitude. To complete the picture, one would like to compare the costs for computer searches with manual searches. A number of preliminary observations need to be made, however. First of all the searcher in a manual system must be quite familiar with the various reference tools, abstract journals, bibliographies and card files which he needs to consult. Secondly, he must be near a library facility which has all of these tools enabling him to undertake the search. Thirdly, even having access to all of the world's published indexes, he could still not be able to perform sophisticated searches in depth, as he could using the computer system described earlier. And last, but not least, manual searches would take much longer. A recent Canadian Study has referred to the economics of manual searching:

> "The Economics Subgroup carried out an extensive user study and determined that scientists, engineers and technological managers spend about 15 per cent of their time searching for appropriate STI. Average cost in salaries, as determined by the Economic Subgroup, amounts to more than $1,800 per year. Applied to the 120,000 such professionals in Canadian industry, salary costs alone for STI search exceed $200 million annually." (3)

17. The table shown in figure 6 is an attempt to estimate the cost of manual searches. Although traditional library cataloguing costs seem to be as high, or even higher than, the preparation of machine-readable input described in paragraph 13, no estimation is made in this table of the value of the catalogued data (4).

Figure 6

Cost of performing manual searches

	per search
a) Data preparation cost	(pro memoria)
b) Staff time to search bibliographies, files, etc. (8 hours at $7/hour)	$ 56.00
c) Staff time to type results (3 hours at $3/hour)	$ 9.00
	$ 65.00

18. The cost estimates given in figure 5 relate to the operations of one documentation service. The cost can be further reduced by envisaging the creation of an information network. An on-line network would have the advantage of enabling participating institutions to cut down on traditional costs in the acquisition and processing of materials by avoiding duplication of work at the input stage. A division of labour would also

---

(3) Science Council of Canada. *Scientific and Technical Information in Canada*, pt. 1, Ottawa, Queen's Printer, 1969. p. 5.

(4) A 1967 estimate of $7.30 unit cost has been made concerning staff costs in the Columbia University Libraries Science Division (Paul J. Fasana, "Processing costs for science monographs in the Columbia University Libraries". *Library resources and technical services*, vol. 11, no. 1, Winter 1967, p. 104). A projection to 1972 of staff costs for the manual system of the Library of Congress gives an estimated unit cost of $6.58 for descriptive cataloguing and $2.91 for subject analysis and classification. (*Automation and the Library of Congress*. Washington, 1963, p. 87.)

enable a much larger information store to be built up, thereby lowering the cost to each participant of data preparation. Furthermore, developmental costs for the computer system could be shared. This would be particularly beneficial for the further improvements in the system that will be necessary to keep apace with technological and other changes. The concluding paragraphs of this paper relate to efforts being currently undertaken to create two such networks.

IV. INFORMATION NETWORKS

19. The International Committee for Social Sciences Documentation has been building up since 1952 a network comprising national centres who submit cataloguing copy for important new publications (books, journal articles, government documents) in four disciplines of the social sciences: sociology, economics, political science and social and cultural anthropology. From the mass of material produced in the world each year, some 25,000-30,000 items are selected and published in the four annual volumes of the *International Bibliography of the Social Sciences*. The cumulative manual store of items published to date in the bibliography comprises some 300,000 items.

20. The ICSSD is presently preparing a bi-lingual thesaurus in economics, which will be tested in an experimental SDI during the course of 1970. In addition a sociology thesaurus is scheduled for completion in late 1970. The existence of these two tools will enable the ICSSD to seek means of recording input data in machine-readable form, opening the door to a number of information services to social scientists.

21. For the past three years, an informal network has been created which links together international organisations active in economic and social development with bilateral aid agencies and other institutions having similar subject interest. Under the patronage of the OECD Development Centre, which coordinates the activities of the network through its Development Enquiry Service, several of the correspondents have been working on the preparation of an indexing vocabulary, the first edition of which was issed in March 1969 (5), as a joint publication of the Deutsche Stiftung für Entwicklungsländer, the Food and Agriculture Organization of the United Nations, the International Committee for Social Sciences Documentation, the International Labour Office and the OECD Development Centre. Since then the indexing vocabulary has been expanded to take into account the requirements of a number of other important international organisations.

22. Considerable work remains to be done, however, to test the vocabulary and properly structure it into a thesaurus. This task assumes increasing importance when one considers the possible use of the *Aligned List of Descriptors* in an on-line inter-agency information network. Co-ordination of the vocabulary development for both this network and the International Committee for Social Sciences Documentation takes place in Paris, under the guidance of Mr. Jean Viet, head of the Service d'echange d'informations scientifiques of the Maison des sciences de l'homme.

23. To proceed from an informal operation into a true information network, it was first necessary to create a common descriptor list acceptable to all participants. The next phase will be to agree on a division of labour to avoid duplication of effort and on a communications format so that machine-readable copy may flow from one operating system into another. Inter-agency co-operation in software development will also be an important factor in accelerating the setting up of such an operational network.

24. The development of this network and the tools needed to make it work represents a rather unique common effort on the part of research institutes and various national and international organisations. It may take another three or five years for the network to become fully operational. When it does, a useful tool will have been created which should provide a means for making a vast storehouse of information available to those social scientists and others who are attempting to come to grips with the pressing economic and social problems of our time.

---

(5) OECD. *Economic and social development aligned list of descriptors.* Paris, 1969, 3 vols. (trilingual edition: English-French-German)

ELECTRONIC RECORDING OF EDUCATIONAL DATA IN THE CANTON OF GENEVA

by

W HUTMACHER

Introduction 113

The Geneva canton schools information system 114

Possible developments 119

Repercussions on school planning 123

Conclusions 125

This paper is available in the original French on application to the Secretariat.
Cette étude est disponible dans sa version originale (français) auprès du Secrétariat.

ref. DECS/Doc (69)19

INTRODUCTION

No doubt under the influence of the acceleration of changes in socio-economic structures and the bewildering increase in the knowledge to be transmitted from generation to generation, modern school systems have been subjected for some decades to outside pressures which, without being basically new, are nonetheless extremely severe. In the years to come, schools will have to transmit more knowledge to a greater number of pupils and provide a larger proportion of them with educational, professional and cultural qualifications of a standard so far restricted to a small minority. It is under these conditions and only relatively recently that schools and their accompanying phenomena have begun to be studied by the empirical methods hitherto restricted to the natural sciences. Progressively the work of economists and sociologists has established an approach which consists basically in viewing the school system like any other system of production in the economy. Current econometric theories have been applied to the analysis of the economic aspects of educational expenditure, of the returns on funds invested, of the output of the school system at the various levels of qualification and of the adjustment of that output to the needs of society.

The resulting insertion of education into the overall system of production, and in particular into the context of economic development, represents in itself a leap forward in the general attitude of industrial societies towards productive activities as well as to school activities. But this new outlook needs to be qualified in two ways.

On the one hand, economic and financial considerations are far from being the only ones which justify striving for higher output from the schools. The need to be educated also makes itself felt in the spheres of politics and culture, so true is it to say that the growing complexity of industrial society requires from each citizen an unprecedented degree of participation in social life.

On the other hand, up to the present the school system as a productive system appears in its various forms as determinate and relatively immutable. The problem of the technical progress and productivity of the school is sometimes envisaged but scarcely analysed and certainly not solved. The problem of the variables which govern - from without or within - the academic progress of pupils has often been studied by sociologists and pedagogues, but usually on sample data valid for a limited period only.

This approach, either too global or too partial, is not surprising in itself. In the majority of countries, there is an almost total lack of sufficiently detailed information to allow a thorough analysis of the internal mechanisms of the school system in its totality. This dearth of statistical data can in turn be explained by various reasons. The ideological connotation of the very notion of measurement where children are concerned - a sphere for qualitative values if ever there was one -is doubtless in part responsible. Perhaps also the outside pressures on the system have never been sufficient to force it to question its own working efficiency. But, on the other hand, the very diversity of the school system and the length of attendance at school are of such a nature as to discourage at the outset any attempt at a global approach. To take account of every aspect, it is indeed necessary, when recording educational phenomena, to use an extremely wide range of criteria over a period which may, for certain categories of pupil, stretch over 15 or 20 years. Finally, the very fact that the school system is an aggregate of diverse types of human behaviour adds the problems inherent in the qualitative nature of the variables and the probabilistic character of the measurements to those arising out of the aforementioned variety of empirical criteria.

But, whatever the reasons for the present situation, one can now observe in all circles having an interest in schools a certain unanimity in favour of striving for "technical progress in schools" and in support of the idea that, as in other fields of activity, a scientific knowledge of how the educational process functions internally will allow lines of action to be defined which may reasonably be expected to improve the output of the school more than proportionally to the sums invested. Experimental pedagogy, based on the discoveries of psychology, is already engaged in such improvements. But it cannot achieve everything alone. The structures of the schools themselves can harbour major obstacles to development which are not the province of pedagogical research in the strict sense. Further, research in sociology and economy has shown that the integration of the education system with other social systems (family, social class, etc.) is much more intimate and complex than at first supposed.

The first problem facing pedagogues, sociologists and economists concerned with the school system consists in assembling a great amount of information on the present state of affairs, with due regard to the existing educational structure as well as to the total length of the school career. This first step is necessary both to obtain a satisfactory description of the system as it worked initially and to measure its evolution, the transformations taking place either "spontaneously" (due to the effect of uncontrolled factors), or under the influence of new structures or methods.

As the social sciences develop new approaches to the study of the school system, they will need to have recourse to more powerful measuring instruments to register and store the mass of data required. Among such instruments, electronic data processing machines will play a considerable role in the future, because of their calculating capacity and above all because they permit the constitution under laboratory conditions of files which reproduce every detail of the school system.

In 1963 the Department of Education of the canton of Geneva gave its sociological research unit the task of compiling an electronic file of the pupils and students of the canton's entire education system, from nursery school to university, public and private.

In an ordinary business firm or an average administrative department such an undertaking would have been justified by the mere relief which generally results from the mechanisation of administrative tasks formerly done by hand. However, without neglecting these aspects, priority was given in the Geneva information system to the collection of scientific data.

This information system is at present still only in the embryonic stage, but it is destined to become ultimately a real "information bank", capable of furnishing rapidly exhaustive information on all educational activities in the canton.

The purpose of this report is to describe the present system in its day-to-day operation, to speculate on possible developments and to describe some of the most interesting applications, in particular from the point of view of educational planning.

1. THE GENEVA CANTON SCHOOLS INFORMATION SYSTEM

The compilation of the basic file of all the pupils and students in the canton took three years for reasons which did not only depend on the difficulties inherent in such an operation. Numerous tentative efforts in various directions, the delay in gaining access to the necessary data-processing equipment, together with a certain lack of special training on the part of those responsible for the project contributed a great deal to prolonging the compilation period, which may appear to have been too long.

1.1. *Population*

At present, the file contains records for some 60,000 pupils and students of the following public and private institutions:

- nursery schools
- primary schools
- lower secondary schools
- upper secondary schools (academic)

- upper secondary schools (vocational)
- part-time vocational courses for apprentices
- full-time vocational schools for apprentices
- university

A sample survey, conducted in 1969, gives grounds for thinking that approximately 1,500 full-time pupils at private vocational schools may still not be covered by the information system (principally typing schools, business colleges, etc.).

1.2. *Variables*

The information available on each pupil at present consists of a record of 440 positions. The different variables recorded are:

Personal variables: (1)

- surname, first name *(A)*
- date of birth *(A)*
- sex *(B)*
- father's first name *(A)*

In order to be able to add information concerning a particular pupil from one year to the next, *a pupil's number (B)* is used as well, based on the components of these personal variables.

Social variables:

- nationality *(A,B)*
- religion *(A,B)*
- father's occupation *(A)*
- socio-occupational class *(B)*
- year of arrival in the canton of Geneva *(A)*
- capacity of person responsible *(A)*

Spatial variables:

- full address (street, number, postal number, town or village) *(A,B)*
- commune of residence *(B)*
- situation on the grid plan of the canton *(B)*

Scholastic variables:

- school, class, section, stage of education, year t *(A,B)*
- school, class, section, stage of education, year t-1 *(A,B)*

---

(1) The variables can be recorded in clear *(A)* or in code *(B)* or in both forms *(A,B)*.

- diplomas obtained *(B)*
- teacher's status *(B)*
- teacher's sex *(B)*

(last two entries concern single-teacher classes only).

This information is stored on magnetic tape. It has been obtained directly from the pupils or their parents and has already had several additions made. The method of collection differed according to the age of the pupils, and according to whether the school was public or private. No mention need be made of the difficulties encountered during this initial stage: they were, all things considered, trifling.

1.3. *Up-dating*

On the other hand, keeping such a file up to date requires very special attention. Arrivals, departures and transfers of pupils are frequent throughout the whole school year and, once a year, each individual is at least assigned to a new class. The value of the file would rapidly be reduced to nothing if a good method of up-dating information were not immediately brought into use. In the case of Geneva, changes are processed in two different stages: annually and daily.

1.3.1. *Annual up-dating*

Principles

From one school year to the next it is important:

- to *preserve* all unchanged information (normally all data except the level reached in the school)
- to *re-register* all pupils and students rapidly at their new level.

Procedure

It was further decided to superimpose on this operation procedures for checking the information stored in the system. These operations are at present carried out in the following stages:

(a) At the end of an academic year (June), pupil cards for all individuals recorded in the system are prepared by computer print-out. These cards refurnish to parents or pupils all the information at the system's disposal, except the scholastic variables which are the ones that are going to change. These cards are punched with an arbitrary serial number which, henceforth, will also appear on the pupil's record in the file.

(b) Grouped by classes, these cards are sent to the schools where the class teachers distribute them to their pupils along with the school report.

(c) The pupils (or their parents) are requested:

- to preserve the pupil's card carefully
- to make any necessary corrections to the information printed by the computer
- to bring the card back to their new class teacher at the beginning of the new academic year, no matter which school they attend in the canton.

(d) When the new year begins the teachers receive their own teacher's card. They collect their pupils' cards and, with their own teacher's card on top, send them to the data processing unit. The teacher's card is punched with a number referring to a school code designating the characteristics of the group of pupils (level, section, school, etc.) as well as those of the teacher (where relevant).

(e) As soon as these cards are received by the processing unit all that is needed to constitute the new file for the year is to feed in the numbers of all the pupils in a class together with the number on the teacher's card, then to retrieve from the preceding year's file the original data (theoretically correct) appearing under each pupil's number. The corrections to the original information supplied by pupils or parents are then fed in. Simultaneously the information recorded under "current school year" is transferred to "previous year" and replaced by the details of the new school situation.

This method has in principle been applied for five years with good results . It has been adhered to because the school appears to be the most rapid and effective channel for distributing documents to pupils and because the information system, as conceived at present, "knows" in which class a pupil may be reached at the end of a school year, whereas it would not "know" at the beginning of a new year. Certain correctives need however to be applied to this basic procedure. For one thing university students are registered under a different system, peculiar to the university. Also, new pupils enter the various schools and classes at primary and secondary level who, naturally, did not appear in the previous year's file. Finally, a certain number of pupils lose or damage their pupil's card between June and September (approximately 5 - 7%).

New pupils starting at nursery or primary school (at present 4,000 per year) are enrolled in May. The necessary information is fed into the card index and, just before the school year begins (August), the parents receive a pupil's card by post.

All other pupils who, for one reason or another, do not posses their card on returning to school, fill in a new one. The "census" for the new year begins 10 days after school begins, to allow a certain time for initial transfers and late arrivals. The file is finally constituted three to four weeks later.

1.3.2. *Daily up-dating (in the course of the academic year)*

Every departure, arrival or transfer of a pupil is notified by the school concerned. Recorded in several copies on NCR self-duplicating paper, the notification is immediately distributed to the various services concerned, including the data processing unit. These up-dating operations are carried out daily at the processing unit so that at any given moment the file exactly reflects the situation in the schools.

1.4. *Codes*

In order to be available for use, the information collected may be recorded on magnetic tape, either in clear, or in code, or in both forms. The creation and up-dating of codes is in itself a considerable task. Certain codes remain very constant (sex, nationality, religion for example), others on the contrary change frequently (the school code differs from year to year), but to allow for comparisons at various points in time, they must satisfy minimum criteria of permanence of logical structure over a long period.

It would be tedious here to enter into details of how codes are established; let it suffice to say that they require the greatest care in order to avoid an unnecessarily complicated program (basic reason for using codes) or repeated operations (where codes are too contracted, for example).

1.5. *Case histories file*

Information concerning a given pupil in different school years may be automatically consolidated thanks to the pupil's number, which remains constant because it is based on the personal variables recorded in the file. The combined information thus builds up a file of case histories which, with time, will contain a complete record of the school career of each pupil. This file is now being constituted and may be used equally for individual consultation (to find out a pupil's scholastic record), for gleaning information on groups of pupils (for example an analysis of the school careers of various cohorts of the pupil population), and finally and most important, for studying the flow of pupils through the school system as a whole.

1.6. *Applications*

The possible applications of such a file are numerous. Our task here is not to list them all but rather to show their variety. Moreover, new possibilities of application are continually being added to the school information system.

1.6.1. *Administrative applications*

Until 1967-68, class teachers were obliged, at the start of the new school year, to write out 5 or 6 lists of the names of pupils in their class. Apart from being written by hand, and therefore sometimes difficult to read, there were still not enough lists, and numerous type-written copies had to be made. These lists are now produced by computer with mechanical speed and precision. It should further be noted that since class teachers have been relieved of this task, thanks to the electronic file, liaison between them and the information system has considerably improved, which, in turn, is good for the accuracy of the information and the speed with which it is transmitted. The data processing unit at present issues between 13 and 15 name lists, or other working documents, for all the departments concerned with school work or the school population (inspectors, teachers, school health services, chaplains, school insurance, vocational training bureaux, etc.).

The printing of labels with pupils' or their parents' addresses or the addressing of documents to be posted direct to families is also done currently on request throughout the year.

In other cases, specialised card indexes for the collection of information of a pedagogical nature (e.g. common examinations) or concerning health (anthropometric measurements, TB checks, etc.) are extracted from the general index. The users are able to profit in this case from information already collected and coded which can either serve for addressing documents (name, address), or be brought into play as control variables when newly collected information is being processed (nationality, sex, social background, etc.). These administrative operations allow a gain in time and in precision, and thus help to offset the cost of the computer file; it is aimed to develop them but they very often necessitate a reorganisation of the services making use of the computer.

1.6.2. *Statistical documentation*

The most important advantages of the information system probably derive from the avoidance of losses thanks to the good knowledge of the present situation and the forecasting of the future which are possible with a complete file of pupils. Such a system in fact allows both varied and detailed statistical documentation to be established rapidly, without it being necessary on each new occasion "to conduct an enquiry". The routine program of school statistics in Geneva includes at present some 100 separate breakdowns a year. But each year new and more refined questions are posed, which can be answered in most cases by statistics derived from the file. This adaptability is indispensable in a system as complex as the educational one where, in addition, the answer to one question often gives rise to others.

From the statisticians' point of view the existence of an electronic file is advantageous in particular because the information is gleaned at its source and because, as it is also being used for administrative purposes, it is much more closely checked, and therefore more accurate. From the point of view of questionnaire design, the heads of information may be all the more numerous because it is not necessary to collect and re-code for each annual operation those data which remain constant over the years; they are simply retained and transferred from one year's file to the next.

The business of forecasting pupil population is only indirectly bound up with the file since it can very well be carried out by other statistical methods. But it is indisputable that the quality of the statistics improves with the use of an electronic filing system, both from the point of view of the reliability of the information and that of the variety of the data heads. Thanks to the accuracy and variety of current statistics, it has been possible to work out a simple method of forecasting population which has played an important rôle in the operation of the Geneva education system over the past years. Indeed, after having been numbered for long among the regions of Europe with the lowest birth rate, the canton of Geneva became from 1953 to 1964 the scene of a marked rise in birth rate, thanks to the rejuvenation of the population by immigration from abroad. It was possible to forecast the repercussions of this "population bulge" throughout the various types of school by 1962. From these statistics, estimates of requirements in school premises and teachers can generally be derived directly. Nevertheless, the assumptions that must be made for the purposes of such forecasts are sometimes precarious and need to be reviewed regularly. At present, forecasts of pupil population are made annually on the basis of new observations and published at the same time as school statistics.

Similarly, the siting of new schools according to the place of residence of potential pupils was recently studied in the case of secondary schools. This forecasting is another problem which may be solved, partially at least, thanks to the pupils file. We have in fact at our disposal the addresses of those pupils likely to be attending upper secondary schools in 1980 (pupils at present aged from 6 to 7); a study of the distribution of their homes on a map of the canton then makes it possible, at an initial stage, to visualise the foreseeable shifts of density within the boundaries of the canton (in particular a centrifugal movement from the city towards the periphery). This first study has led in turn to a change in the file involving the inclusion (at present under way) of a code positioning pupils on a grid map of the territory, where each square represents an area of 100 x 100 metres. This new variable will allow closer calculations to be made, based on the distances to be covered by pupils and the means of transport at their disposal.

As stated, the list of these practical applications is not exhaustive. In this field, new tasks are added every year, either on the basis of information already available or by virtue of new developments in the system.

2. POSSIBLE DEVELOPMENTS

In its present form the information system constitutes the nucleus of a "school information bank" which has to be extended by stages in different directions.

2.1. *Extension of the population covered*

Certain services of the department of education are not specially concerned with school pupils or students, but more generally with minors; these are principally medical, social, medico-pedagogical and legal services. The reciprocal influences between the social field of action and that of the schools are so far-reaching that consolidation of the information from both sectors would be justified for scientific reasons alone. It may be added that, apart from technical data, the information to be stored is identical for a vast number of individuals (all those who are at once pupils and are undergoing treatment for social, psychological, medical or judicial reasons). The only rational storage system in this type of situation is thus a central index.

Sooner or later, the index should therefore cover all minors from 0 - 20 years old in the canton (total population approximately 80,000), plus pupils or students over 20 years of age. The attainment of this aim still comes up against two obstacles:

(a) Since pupils and students belong to an educational establishment, the information concerning them can be collected by the administration of their establishment. But the same cannot be said of children from 0 - 5 years old and adolescents from 15 to 20 not attending school who, for their part, can be contacted initially only by the local census authorities. External liaison with these authorities is therefore indispensable and is obviously best established by means of the computer. The constitution of a machine-readable general file of all the inhabitants of the canton is already in preparation.

(b) The problems of circulating information between the peripheral services and the information bank (organisation) and those concerning the nature and structure of the information to be circulated (analysis) require very thorough study; for the school system alone the task is already considerable, but there is such a variety of concepts to be studied in the fields of medicine, social psychology and law, that the processing will demand a much higher investment per job.

2.2. *Extension of the information processed*

The reasons invoked to justify extending the population covered may also be used in favour of increasing the quantity of information processed: anthropometric, medical, social, legal data, etc.

For the study of the school system, a range of supplementary information of a pedagogical and sociological order will eventually become essential. Thus the recording of pupils' annual marks would permit interesting docimological analyses, as well as making possible statistically the detection of those subjects which most frequently cause pupils at a given stage to repeat a year. But it will probably be necessary to take a further step and organise school examinations common to a clearly defined group of pupils (one stage of education for example). Properly conceived and standardised for a large number of pupils such examinations would allow a global diagnosis to be made at the level of the school curriculum in general (for example notions which are often difficult to assimilate) as well as at levels closer to the classroom, for example the class group or even the individual pupil. Further, a comparison of results obtained at an examination at the beginning of the school year and one at the end would provide a yardstick of the knowledge acquired by a certain group over the period under study. Comparison of the results obtained at an end-of-year examination and one at the start of the following year would on the other hand reveal facts about losses due to the holidays.

The working out and introducing of such systems are obviously the particular province of education experts and will not be gone into here but, apart from the problem of devising examinations, stress must be laid on the degree to which such work depends on a speedy system of correction and interpretation of results. Considerable advantage will no doubt be taken of optical reader processes in this sphere as well as of the processing and interpretation capacity of computer programs.

2.3. *Extension of the services rendered by the system*

The extra return on capital invested represented by each additional application of the data stored is a primary reason for seeking to make maximum use of computer programs for administrative and technical tasks in the management of the education system. Such changes do not always - perhaps even not often - result in a reduction in departmental staff. Indeed, departments are often under-staffed. At the most, one can hope with the help of computer programs to relieve existing staff of routine tasks, to concentrate their attention on work so far neglected or not even embarked upon. It is well-known how readily organisations tend to deal with urgent, that is to say current, business, and neglect jobs that "can wait", among which are frequently balance-sheets, forecasts, detailed working out of aims and long-term planning.

The services rendered by the information system should not be extended in a haphazard way by superimposing new tasks on the old; here again and perhaps above all, a general development plan is necessary if one wishes to avoid simply executing by computer work formerly done by hand. Attention to this problem leads to a closer analysis of the principles guiding the development of a school information system.

2.4. *Principles of development*

The major objective consists in developing a general information system containing the necessary data and adequate for a full knowledge of the school system and for those administrative applications which lend themselves to mechanisation.

As far as possible the information system should collect its information once, when and where it is produced, and in machine-readable form, in order to avoid delays and in particular communication relays (coding, transcribing, punching, etc.). The quality of the information collected by a department is its own responsibility. It must guarantee that the information it feeds into the system is full and accurate. The analysis of the problem, programming and program maintenance, are entrusted to a special staff who, during the whole preparatory period, collaborate directly with the client department.

The tasks required of the information system will be added to after a thorough study of the possibilities and needs of the services interested in its applications as well as of all educational and social services. This principle brings us directly to the overall organisation of the system.

2.5. *Organisation*

In the present context we will understand by "organisation" all the studies concerning the generation and transmission of information inside a job system. Everyone who has written on the problem of introducing electronic data processing methods into administrative services insists on the need for a very thorough initial study of the system to be mechanised. It is often here rather than in the computer jobs themselves that the most important gains in time and efficiency can be made. Administrative services, perhaps more than others, have a tendency to evolve by constantly adding new tasks to old: new schemes, new files, extra staff. The work involved in the school information service could, in such a context, simply be added to existing work or else manual tasks might simply be taken over by a computer. The benefit would in this case be limited to the availability of information processable by computer program (which would certainly satisfy the scientific aim mentioned above) but in the field of administration proper, the gains would be negligible or non-existent.

When an information system is worked out, it is necessary on the contrary to re-examine all the problems in a new light and seek as far as possible to integrate similar jobs, processing the same array or sub-array of data, so as to reduce to a minimum the number of jobs and the quantity of information required. This rethinking of school administration cannot fail to bring about many and often basic changes in work routines and a certain standardisation of organisation and methods in the various schools.

But above all, it must be realised that the reorganisation implies a change in outlook and therefore cannot take effect without the active participation of the persons directly concerned, whose co-operation is the fundamental condition of success. After patient and minute study by all interested parties, reviewing every action with all its underlying principles, a reorganisation plan should be drawn up, which, apart from immediate needs, will also have sought to take account of the long-term development of needs and attitudes. Such a plan should be made widely known and submitted for criticism and suggestions from staff at every level and, after revision, adopted as a working basis. It will specify various levels of responsibility and care will be taken to prevent departments or their staff from "abdicating" their special skills in favour of the planning staff. The responsibilities of the various departments will remain unchanged in the overall plan (except where two departments are amalgamated);

only the working methods will change. This task of reorganisation is particularly difficult in a school system, because human beings are the main element, because it is rarely possible to justify the reorganisation by the performance of material media (for example machines with faster output rates) and because the efficiency of school systems is measured according to complex criteria which are often vague and always imbued with subjectivity.

2.4. *Other information systems*

Other aspects of the educational system call for exploration by the same methods, in particular: staffing, premises, equipment and furniture and the financial and budgeting systems.

Teaching staff

Judged by economic criteria, the school is a production unit with a very high staff density: between 85 and 95 per cent of expenditure is on salaries. For some years the difficulties of recruiting teaching staff have become greater, which has contributed moreover in no small way to the growing awareness within the school system of its dependence on the staff variable. It is therefore necessary for all sorts of reasons to study the structure of teaching staff and its transformation. The increased preponderance of women, ageing, qualifications, in-service teacher training, are all factors that influence the present educational system and, owing to staff permanence coefficients, are likely to transform the system more or less gradually in the future. The preparation of a teaching staff information system employing the methods of the pupil system is at present under preparation in Geneva. As with the pupil system, it must at the same time meet the needs of administration (salaries, budget control, appointments, etc.) and facilitate the empirical analysis of the main transformation parameters which will allow future trends to be reckoned.

Premises, equipment, supplies and furniture

In an education system, there rarely exists a thorough and permanent inventory of the material factors of production. There is therefore usually no study of the extent of their use and their length of life. This may depend partially on the small proportion of the cost of education represented by these factors. But, in economic studies of the cost of education and in spite of the preponderance of the staff factor, these variables should figure under the headings *Maintenance* and *Amortization* and not, as is the case today, only in the form of *new purchases*. It is true that this more economic and financial approach to the education system has emerged only relatively recently. However, by the very extent to which it promises to provide more definite and more refined estimates of the directions in which the education system is evolving, it will no doubt be adopted more and more in years to come, particularly in the field of finance, accounts and budgeting.

Financial and budgetary systems

Accountancy and finance generally figure among the most highly organised services of educational administration, but there again, the financial aspect should be complemented by an economic approach in the form of a study of the cost of education per pupil or student at the various levels. The value of precise information on this subject appears obvious for analysing the present educational system as well as for forecasting the budget. The comparison between the different costs of education per pupil taught at the various levels as well as between the different cost structures (staff, supplies, premises, etc.) provides some extremely interesting facts about the present running of schools. In addition, the budgetary repercussions of a population bulge like the one which the Geneva school system is experiencing at present will be easier to foresee if one can work with pupil costs differentiated according to level of education. It can then be seen that, all other things being equal, the growth of the education budget over a period of time is more than proportional to the growth in the number of pupils because of the continued presence of pupils in the system over a number of years and because the unit cost of education is higher at the upper levels.

These considerations on the possible developments of our global information system in the field of education do not of course claim to be exhaustive, but they at least show the degree of complexity of the overall conception on the one hand, and on the other, because of the diversity of possible approaches, the necessity of close interdisciplinary collaboration.

While conserving their logical coherence, the different information systems and their extensions will no doubt have to be built up by stages. But what is more, the new knowledge arising out of this methodological approach will in the future transform both the school system and the information system. New problems will doubtless have to be faced then, hence the fundamental necessity for flexibility at every level.

3. REPERCUSSIONS ON SCHOOL PLANNING

All planning takes for granted sound knowledge of the basic situation acquired with the aid of meaningful parameters, an estimate of the "spontaneous" evolution tendencies of these different parameters and finally an education policy that weighs up both the present state of affairs and the spontaneous tendencies, defines the changes to be aimed at and indicates how they can be brought about.

It has already been shown how, by the electronic information system, the statistical information required to estimate present day parameters can be secured more rapidly, regularly and accurately.

Repeated observations of the same kind over a period of time give an idea of the spontaneous evolution of these parameters: that is to say, make it possible to forecast. A good estimate of probable future development trends is obviously necessary in every field: pupils, teachers, present use of premises and equipment and new requirements, budget, etc. However, whatever the general line of approach, the estimate of future numbers of pupils at the different levels is the basis of all other calculations.

Now, in this sphere, new problems have recently appeared in the canton of Geneva. Present forecasting of future pupil population in the canton relies principally on demographic data which are transformed into school population figures by the application of a school attendance rate coefficient. (2) For want of satisfactory chronological series, school attendance rates are taken as constant in the present model, which means that no modification of the output parameters of the school system is envisaged. Checks on these forecasts show that they systematically under-estimate the number of pupils actually found, in particular in higher secondary schools. This is probably due to the fact that the school system is changing, and also that there is a growing propensity to go on to higher secondary education. In support of the first hypothesis, one may point in particular to the gradual but steady decrease in the proportion of retarded pupils at all levels of the school system since 1960. Now, in the present school system, this means that a greater number of pupils complete their schooling with success and can as a result opt to prolong their studies.

To obtain a clearer grasp of the internal transformations of the educational system, it is intended in future to rely on parameters from the system itself in the form of a promotion rate. (3) The promotion rate can be calculated from the case histories file mentioned above which, for each pupil, shows his situation in the school in different years. The promotion rates thus calculated form an "input - output" matrix in which the demographic variables appear in the form of lines and columns in relation with outside populations. These outside populations may, in addition, be broken down on the one hand by whether they consist of pupils recruited inside or outside the territory under observation and, on the other hand, by pupils who leave the education system with or without a qualification and whether or not they remain within the territory under study. The analogy between this method of apprehending the school system and the model worked out by H.P. Widmaier is obvious. However, it will not be possible, during the initial phase at least, to bring external economic variables into play (for example manpower requirements) because the canton of Geneva represents too small an economic unit and because economic variables are measured in Switzerland, when at all, at the national level of all cantons combined.

---

(2) School attendance rate = the proportion of residents of a given age attending school at a given level.

(3) Promotion rate = the proportion of pupils at stage n, year t, who move up to stage n+1, year t+1.

Nevertheless, the "input - output" model presents considerable progress from the point of view of forecast methods, since these now rely on the parameters of the school system itself in which it is known future changes are likely to be considerable. At a later stage, failing economic data, it will be possible to add to this model at least the budgetary and financial imperatives, after which it will serve for simulating various possible policies, that is, for carrying out "in the laboratory" experiments which could scarcely be envisaged "in real life" in an education system.

It is already intended to modify somewhat the model described above by studying separately the promotion rates of boys and girls, and of different social and occupational groups. The canton of Geneva finds itself in fact confronted with a situation whose development is difficult to foresee:

- the demographic increase is due mainly to the recent influx of young immigrants from abroad. Whereas they represented only 14% of the births in 1952, the children of immigrants now form half of each new generation.

- This foreign population is largely made up of unskilled workers and it is a recognised fact that the children of this social group do less well than others at primary school and their chances of reaching higher education are far inferior to those of children from more well-to-do families.

- It may therefore turn out that the numbers in higher secondary schools are only slightly affected by the population bulge whereas vocational schools and apprentice training courses may feel the side-effects more than proportionally.

But it is also possible that the educational behaviour of the chidren of unskilled workers may alter, either under the effect of the measures to democratise education which have recently come into force, or where foreign childen are concerned because their parents belong to an immigrant population (on the assumption that mobile populations are more dynamic and ambitious).

The creation of three different promotion rate matrices for three different strata will make it possible to analyse and follow up this problem, which is obviously of capital importance for the future of schools in Geneva and for the reruitment of an ever-growing proportion of highly qualified manpower.

CONCLUSIONS

This report contains many gaps, some of which are due to the limits of the study itself; other and more fundamental omissions are due to the fact that the system in operation in Geneva is still largely experimental. It was moreover not intended to treat the subject exhaustively. It was rather a question of giving a glimpse of the general aims of the undertaking, to explain its essential technical details and above all to show how much it depends in turn on inter-disciplinary co-operation and long-term planning.

However attractive and fruitful they may be, more or less "mathematised" models of the school system should not obscure the fact that they are only models, with all the methodological limitations that implies, and that other models are possible and will no doubt be produced in the future, by very reason of the changing attitudes of those responsible for running the school system, as well as of the new needs of the social system in general.

NOTE - a detailed Appendix to this paper is included with the original French version which is available on application to the Secretariat.

Notes on Authors

COATES E J

Editor of *British Technology Index* since 1962; formerly of British National Bibliography. Author of *Subject Catalogues*, *British Catalogue of Music Classification*, and various papers on classification, notation, vocabulary control, computer aided indexing, and switching languages for linking indexing systems.

HUTMACHER Walo

Born 1932. Graduated in sociology at the University of Geneva; Assistant to Professor Girod; Director of an institution for maladjusted and delinquent adolescents; freelance sociological researcher; Director of the Sociological Research Service, Geneva.

PESTON Professor M H

Read economics at the London School of Economics and subsequently held a Fellowship at Princeton University for two years. On the staff of the London School of Economics from 1956 until 1965. He was also an economic consultant to the Treasury until 1964, and subsequently a part-time consultant to the Ministry of Defence. He is now head of the new department of economics at Queen Mary College, London.

PRICE A G

Born in 1934. Gained an Honours Degree in Mathematics at Oxford in 1955. Joined ICI Ltd. in 1957 and worked on commercial data processing by computer until 1961, when he joined English Electric (now ICL). He was involved in the programming of an ALGOL compiler, and in implementation and design of computer operating systems. In 1967 took charge of the Documentation Processing Centre, a part of the Office for Scientific and Technical Information, to direct the provision of computing facilities for experimental projects supported by OSTI and research into computer methods applicable to information processing.

STRÄTER Hans H

Born 1938. Studied German and English literature and pedagogics; PhD. Head of the Documentation Section, Pädagogisches Zentrum, Berlin.

THOMPSON George K

Born 1929. MS, Western Reserve University Library School 1951. 1951/1952 - Reference Librarian, New York Public Library; 1952/1961 - Documents Librarian, United Nations Library, Geneva; 1961 to date - ILO, at present Chief, Central Library and Documentation Branch.